

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-286680

(43)Date of publication of application : 01.11.1996

(51)Int.Cl.

G10K 15/00

G08T 7/00

G10K 15/12

G10L 3/00

G10L 3/00

H04N 7/18

(21)Application number : 07-217859

(71)Applicant : TAKENAKA KOMUTEN CO LTD

(22)Date of filing : 25.08.1995

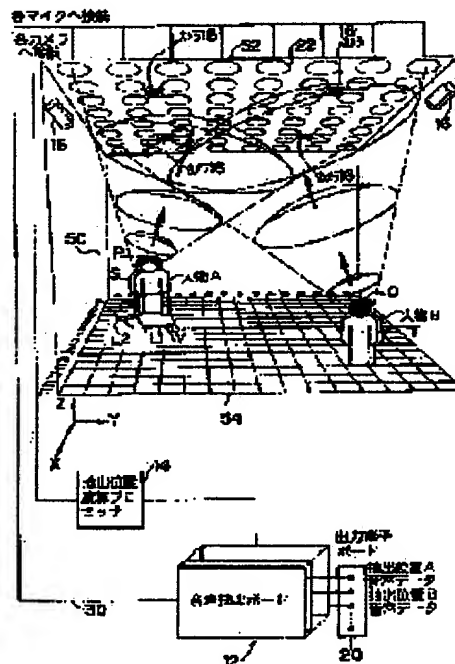
(72)Inventor : HARAKAWA KENICHI
IGAWA NORIO
SOU DAISAKU
KAGEYAMA KENJI
NAKAJIMA TATSUMI
WAKINAKA YOSHITAKA
YAMADA SUKEO
UNNO KENICHI
MURAI NOBUYOSHI

(30)Priority

Priority number : 07 29185 Priority date : 17.02.1995 Priority country : JP

(54) SOUND EXTRACTING DEVICE

(57)Abstract:

PURPOSE: To detect a position of an object whose position is uncertainty and to extract a sound generated by the object based on its position.**CONSTITUTION:** When a person A moves, the inside of a room 50 is continuously photographed by plural television cameras 16, a sound extracting device recognizes its position based on the picture information regarding a region provided with specific feature quantity in a human head of which almost surface is covered by hair and which has much black part and almost spherical shape as a region corresponding to a head part P. Next, sound signals of each collected sound are delayed by a delay time in accordance with a distance between positions of each microphone 22 and a position of a head part P so that a voice part of the person A in a voice signal of a collected sound collected by plural microphones 22 is synchronized along time base. And only a voice of the person A is extracted by adding and averaging a delayed voice signal.

LEGAL STATUS

[Date of request for examination]

07.06.2001

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

STANDARD FORM NO. 64

OFFICE OF THE SECRETARY OF DEFENSE
WASHINGTON, D.C. 20301

1. TITLE
2. NUMBER
3. DATE
4. AUTHOR
5. SUBJECT

CLASSIFICATION

1. COPIES OF THIS REPORT

2. COPIES OF THIS REPORT

THIS PAGE BLANK (USPTO)

3. COPIES OF THIS REPORT

4. COPIES OF THIS REPORT

1. TITLE
2. NUMBER
3. DATE
4. AUTHOR
5. SUBJECT

THIS PAGE BLANK (USPTO)

CLASSIFICATION

1. COPIES OF THIS REPORT

2. COPIES OF THIS REPORT

CLASSIFICATION

1. TITLE
2. NUMBER
3. DATE
4. AUTHOR
5. SUBJECT

CLASSIFICATION

1. COPIES OF THIS REPORT

CLASSIFICATION

1. COPIES OF THIS REPORT

CLASSIFICATION

1. COPIES OF THIS REPORT

CLASSIFICATION

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-286680

(43) 公開日 平成8年(1996)11月1日

(51) Int.Cl. ⁸	識別記号	庁内整理番号	F I	技術表示箇所
G 1 0 K 15/00			G 1 0 K 15/00	L
G 0 6 T 7/00			G 1 0 L 3/00	5 1 1
G 1 0 K 15/12				5 5 1 G
G 1 0 L 3/00	5 1 1		H 0 4 N 7/18	D
	5 5 1			K

審査請求 未請求 請求項の数22 O L (全 33 頁) 最終頁に続く

(21) 出願番号 特願平7-217859

(22) 出願日 平成7年(1995)8月25日

(31) 優先権主張番号 特願平7-29185

(32) 優先日 平7(1995)2月17日

(33) 優先権主張国 日本 (J P)

(71) 出願人 000003621

株式会社竹中工務店

大阪府大阪市中央区本町4丁目1番13号

(72) 発明者 原川 健一

千葉県印旛郡印西町大塚1-5 株式会社

竹中工務店技術研究所内

(72) 発明者 井川 憲男

千葉県印旛郡印西町大塚1-5 株式会社

竹中工務店技術研究所内

(72) 発明者 荘 大作

千葉県印旛郡印西町大塚1-5 株式会社

竹中工務店技術研究所内

(74) 代理人 弁理士 中島 淳 (外4名)

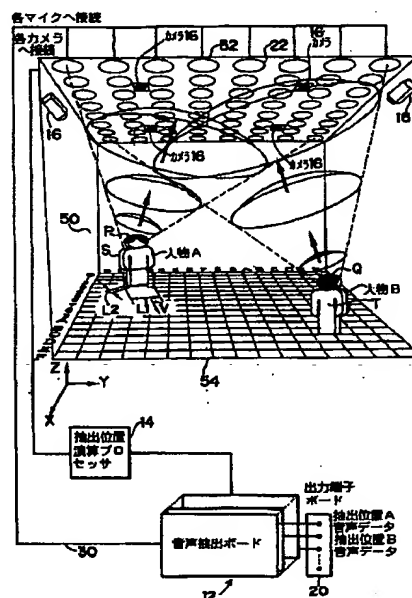
最終頁に続く

(54) 【発明の名称】 音抽出装置

(57) 【要約】

【課題】 位置が不確定であるオブジェクトの位置を検出し且つその位置に基づいてオブジェクトが発する音を抽出する。

【解決手段】 人物Aが移動した場合、複数のテレビカメラ16により部屋50の内部が継続して撮影され、音抽出装置は、それらの画像情報に基づいて、表面の多くが毛髪で覆われており黒色部分が多い略球状であるという人間の頭部に特有の特徴量を備えた領域を頭部Pに相当する領域とみなし、その位置を認識する。次に、複数のマイク22の各々により取られた採取音の音声信号における人物Aの声の部分の部分を時間軸に沿って同期させるべく、各採取音の音声信号を、各マイク22と頭部Pの位置との距離に応じた遅延時間だけ遅延操作する。そして、遅延操作した音声信号を加算平均することにより、人物Aの声のみを抽出する。



THIS PAGE BLANK (USPTO)

【特許請求の範囲】

【請求項1】 音源としてのオブジェクトを含む領域を撮影する撮影手段と、

前記撮影手段により撮影された領域の画像情報よりオブジェクトの位置を認識する画像認識手段と、

予め定められた位置に配置され前記オブジェクトが発する音を採取する複数のマイクロフォンと、

前記複数のマイクロフォンの各々により採取された採取音の時系列データのうち複数の採取音の時系列データを選択し、選択した採取音の時系列データを、前記画像認識手段により認識されたオブジェクトの位置及び該選択した採取音を採取したマイクロフォンの位置に基づいて、オブジェクトが発する音が同期するようにシフトし、シフトした採取音の時系列データを平均することにより、オブジェクトが発する音を抽出する抽出手段と、を有する音抽出装置。

【請求項2】 前記画像認識手段は、オブジェクトを含む領域の画像情報より該オブジェクトが発する方向も認識し、さらにオブジェクトの位置及び該オブジェクトが発する方向に基づいて、オブジェクトが発する音を良好に抽出可能な位置をオブジェクトの位置と改めて認識する、ことを特徴とする請求項1記載の音抽出装置。

【請求項3】 オブジェクトが移動する場合、前記撮影手段はオブジェクトの移動に従って該オブジェクトを含む領域を撮影する、

ことを特徴とする請求項1記載の音抽出装置。

【請求項4】 オブジェクトが複数存在する場合、前記撮影手段は複数のオブジェクトを含む領域を撮影し、前記画像認識手段は撮影された領域の画像情報より複数のオブジェクトの各々の位置を認識し、前記抽出手段は該複数のオブジェクトの各々からの音を抽出する、

ことを特徴とする請求項1記載の音抽出装置。

【請求項5】 前記オブジェクト及び前記複数のマイクロフォンを含む領域において音速及び音伝搬経路の少なくとも一方に影響を及ぼす要因とされる音響環境状態を検出する検出手段を更に備え、

前記抽出手段は、前記検出手段によって検出された音響環境状態が変化した場合に、変化した音響環境状態に基づいて、前記採取音の時系列データのシフトを補正することを特徴とする請求項1記載の音抽出装置。

【請求項6】 前記抽出手段は、高音域の指向性に関する情報に基づき、高音域の採取音の時系列データを重み付けして平均することを特徴とする請求項1記載の音抽出装置。

【請求項7】 前記画像認識手段は、オブジェクトを含む領域の画像情報より該オブジェクトが発する方向、該オブジェクトの周辺に位置する音の反射面の位置及び向きもさらに認識し、

前記抽出手段は、前記選択した採取音を採取したマイクロフォンの位置、オブジェクトの位置、該オブジェクトが発する方向、及び前記反射面の位置並びに向きに基づいて、前記選択した採取音の時系列データを、オブジェクトからの直接音又は前記反射面で反射された反射音の何れか一方が同期するようにシフトする、ことを特徴とする請求項1記載の音抽出装置。

【請求項8】 前記抽出手段は、前記複数のマイクロフォンの各々により採取された採取音の時系列データのうち、前記オブジェクトの位置から所定距離以上離れて位置するマイクロフォンによって採取された採取音の時系列データを、選択の対象から除外することを特徴とする請求項1記載の音抽出装置。

【請求項9】 前記抽出手段により抽出されたオブジェクトが発する音を所定の音声認識装置へ出力する出力手段を更に備えたことを特徴とする請求項1記載の音抽出装置。

【請求項10】 予め定められた位置に配置された広角固定焦点レンズを備え、音源としてのオブジェクトを含む領域を撮影する撮影手段と、

前記撮影手段により撮影された領域の画像情報よりオブジェクトの位置を認識する画像認識手段と、

予め定められた位置に配置され前記オブジェクトが発する音を採取する複数のマイクロフォンと、

前記複数のマイクロフォンの各々により採取された採取音の時系列データのうち複数の採取音の時系列データを選択し、選択した採取音の時系列データを、前記画像認識手段により認識されたオブジェクトの位置及び該選択した採取音を採取したマイクロフォンの位置に基づいて、オブジェクトが発する音が同期するようにシフトし、シフトした採取音の時系列データを平均することにより、オブジェクトが発する音を抽出する抽出手段と、を有する音抽出装置。

【請求項11】 前記撮影手段は複数設けられており、各撮影手段は、前記広角固定焦点レンズによる結像点に配置されたエリアセンサを更に備え、

前記画像認識手段は、該複数の撮影手段により撮影された各々異なる撮影情報を処理してオブジェクトの形状を認識する形状認識手段と、前記形状認識手段により認識されたオブジェクトの3次元座標を演算する3次元座標演算手段と、を含んで構成された、

ことを特徴とする請求項10記載の音抽出装置。

【請求項12】 前記形状認識手段は、前記複数の撮影手段により撮影された各々異なる画像情報に基づいて、3次元空間をX軸、Y軸及びZ軸の各方向に沿って仮想的に細分割することにより得られる多数の立方体状の微小空間のうち、オブジェクトが占有する微小空間により形成される領域を求めることによりオブジェクトの形状を認識することを特徴とする請求項11記載の音抽出装置。

【請求項13】 前記形状認識手段は、前記複数の撮影手段により撮影された各々異なる画像情報に基づいて、3次元空間をX軸、Y軸及びZ軸の各方向に沿って仮想的に細分割することにより得られる多数の立方体状の微小空間のうち、各撮影手段からオブジェクトを投影する視野角内に含まれる微小空間をそれぞれ抽出し、抽出した微小空間の全てに含まれる微小空間により形成される領域を求めることによりオブジェクトの形状を認識することを特徴とする請求項1記載の音抽出装置。

【請求項14】 前記撮影手段は複数設けられており、各撮影手段は、前記広角固定焦点レンズによる結像点に配置されたエリアセンサを更に備え、前記画像認識手段は、各撮影手段のエリアセンサ上に結像された2次元座標を取得し、該取得した複数の2次元座標に基づいてオブジェクトの位置を認識することを特徴とする請求項10記載の音抽出装置。

【請求項15】 予め定められた位置に配置された広角固定焦点レンズ及び該レンズによる結像点に配設されたエリアセンサを備え、音源としてのオブジェクトを含む領域を撮影する撮影手段と、前記撮影手段の近傍に配置され、前記エリアセンサ上に結像するようにオブジェクトの像を反射する反射手段と、

前記反射手段により反射されて前記エリアセンサ上に結像されたオブジェクト像、及び前記反射手段により反射されることなく前記エリアセンサ上に結像されたオブジェクト像の各々の該エリアセンサ上における2次元座標を取得し、該取得した複数の2次元座標に基づいてオブジェクトの3次元座標を演算することにより、オブジェクトの位置を認識する画像認識手段と、予め定められた位置に配置され前記オブジェクトが発する音を採取する複数のマイクロフォンと、前記複数のマイクロフォンの各々により採取された採取音の時系列データのうち複数の採取音の時系列データを選択し、選択した採取音の時系列データを、前記画像認識手段により認識されたオブジェクトの位置及び該選択した採取音を採取したマイクロフォンの位置に基づいて、オブジェクトが発する音が同期するようにシフトし、シフトした採取音の時系列データを平均することにより、オブジェクトが発する音を抽出する抽出手段と、を有する音抽出装置。

【請求項16】 前記画像認識手段は、オブジェクトを含む領域の画像情報より該オブジェクトが発する方向も認識し、さらにオブジェクトの位置及び該オブジェクトが発する方向に基づいて、オブジェクトが発する音を良好に抽出可能な位置をオブジェクトの位置と改めて認識する、ことを特徴とする請求項10乃至15の何れか1項に記載の音抽出装置。

【請求項17】 オブジェクトが複数存在する場合、前

記撮影手段は複数のオブジェクトを含む領域を撮影し、前記画像認識手段は撮影された領域の画像情報より複数のオブジェクトの各々の位置を認識し、

前記抽出手段は該複数のオブジェクトの各々からの音を抽出する、

ことを特徴とする請求項10乃至15の何れか1項に記載の音抽出装置。

【請求項18】 前記オブジェクト及び前記複数のマイクロフォンを含む領域において音速及び音伝搬経路の少なくとも一方に影響を及ぼす要因とされる音響環境状態を検出する検出手段を更に備え、

前記抽出手段は、前記検出手段によって検出された音響環境状態が変化した場合に、変化した音響環境状態に基づいて、前記採取音の時系列データのシフトを補正することを特徴とする請求項10乃至15の何れか1項に記載の音抽出装置。

【請求項19】 前記抽出手段は、高音域の指向性に関する情報に基づき、高音域の採取音の時系列データを重み付けして平均することを特徴とする請求項10乃至15の何れか1項に記載の音抽出装置。

【請求項20】 前記画像認識手段は、オブジェクトを含む領域の画像情報より該オブジェクトが発する方向、該オブジェクトの周辺に位置する音の反射面の位置及び向きもさらに認識し、

前記抽出手段は、前記選択した採取音を採取したマイクロフォンの位置、オブジェクトの位置、該オブジェクトが発する方向、及び前記反射面の位置並びに向きに基づいて、前記選択した採取音の時系列データを、オブジェクトからの直接音又は前記反射面で反射された反射音の何れか一方が同期するようにシフトする、

ことを特徴とする請求項10乃至15の何れか1項に記載の音抽出装置。

【請求項21】 前記抽出手段は、前記複数のマイクロフォンの各々により採取された採取音の時系列データのうち、前記オブジェクトの位置から所定距離以上離れて位置するマイクロフォンによって採取された採取音の時系列データを、選択の対象から除外することを特徴とする請求項10乃至15の何れか1項に記載の音抽出装置。

【請求項22】 前記抽出手段により抽出されたオブジェクトが発する音を所定の音声認識装置へ出力する出力手段を更に備えたことを特徴とする請求項10乃至15の何れか1項に記載の音抽出装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、音抽出装置に係り、より詳しくは、オブジェクト（本発明では、音を発すると予想される人間又は物体）が発する音を抽出する音抽出装置に関する。

【0002】

【従来の技術及び発明が解決しようとする課題】 従来、建物や橋等の建造物の老朽化の度合いを検査する場合、該建造物の所定の部位から発せられるきしみ音等を主にシミュレーションにより算出することにより老朽化の度合いを検査していた。しかし前記シミュレーションにより算出される値はあくまでも予測値であるので、より厳密な検査を行うためには該建造物の所定の部位から発せられる実際のきしみ音を周辺の雑音と弁別して抽出することが望まれていた。

【0003】一方、上記に関連して、複数のマイクロフォンにより集音された音の音声信号を同じ時間軸上に重ね、それらの音の各々に対しマイクロフォンと目的とする音源との距離に応じて適切な遅延操作を行ってから加算平均することにより、目的とする音源から発せられた音のみを抽出する技術が知られている。またこの技術を撮影と録音とを同時に行う手持ち式のビデオカメラに適用し、撮影における被写体から発せられた音を抽出することにより、被写体の映像の焦点と録音の焦点とを一致させる技術が知られている（特開平5-308553号公報参照）。

【0004】しかし、上記の手持ち式のビデオカメラに関する技術は、実際には当該ビデオカメラの撮影視野内という狭い領域内に位置する単一の被写体に対してのみ有効であり、手持ち式であるためにビデオカメラに付属したマイクロフォンは少数であり、それらの配置間隔が狭いため、雑音の影響が大きく、臨場感の高い音声を集音することは比較的困難であった。

【0005】ところで、従来より天井に複数のテレビカメラを配置し、該複数のテレビカメラにより撮影した画像情報に基づいて、部屋の中に存在するオブジェクトの位置を検出する画像認識技術が存在する。

【0006】しかしながら、オブジェクトが移動する場合、この移動するオブジェクトに合わせて複数のテレビカメラを動かし且つ焦点調整を行うことにより、当該移動するオブジェクトを撮影していた。このように、テレビカメラを動かし焦点調整を行う必要があるため、オブジェクトを含む画像データを得るまでに遅延時間が発生するという問題点があった。

【0007】本発明は、上記のような事実を考慮して成されたものであり、上記の位置の検出に関する画像認識技術と音抽出技術とを組合せ、位置が不確定であるオブジェクトの位置を検出し且つその位置に基づいてオブジェクトが発する音を抽出することができる音抽出装置を提供することを第1の目的とする。また、より効率的にオブジェクトの位置検出を行った上で、その検出位置に基づいてオブジェクトが発する音を抽出することができる音抽出装置を提供することを第2の目的とする。

【0008】

【課題を解決するための手段】 上記第1の目的を達成するために、請求項1記載の発明は、音源としてのオブ

ジェクトを含む領域を撮影する撮影手段と、前記撮影手段により撮影された領域の画像情報よりオブジェクトの位置を認識する画像認識手段と、予め定められた位置に配置され前記オブジェクトが発する音を採取する複数のマイクロフォンと、前記複数のマイクロフォンの各々により採取された採取音の時系列データのうち複数の採取音の時系列データを選択し、選択した採取音の時系列データを、前記画像認識手段により認識されたオブジェクトの位置及び該選択した採取音を採取したマイクロフォンの位置に基づいて、オブジェクトが発する音が同期するようにシフトし、シフトした採取音の時系列データを平均することにより、オブジェクトが発する音を抽出する抽出手段と、を有することを特徴とする。

【0009】この請求項1記載の発明では、撮影手段が音源としてのオブジェクトを含む領域を撮影し、画像認識手段が撮影手段により撮影された領域の画像情報よりオブジェクトの位置を認識する。

【0010】例えば、図2に示すように天井52に設置された複数のテレビカメラ16の各々により撮影された部屋50の中の画像情報より以下のようにして、オブジェクトとしての対象人物Aの頭部Pの位置を認識する。即ち、画像情報より、表面の多くが毛髪で覆われており黒色部分が多い略球状である等のように人間の頭部に特有の特徴量を備えた領域を、対象人物Aの頭部Pに相当する領域として抽出する。そして、抽出した頭部Pに相当する領域が、部屋50を矢印X方向、矢印Y方向、矢印Z方向の各々に沿って仮想的に等分割することにより得られる多数の直方体状の領域のうち何れの領域に相当するかを認識する。

【0011】一方、オブジェクトが発した音は、予め定められた位置に配置された複数のマイクロフォン（以下、マイクと略称する）により採取される。例えば、図2に示すように2人の人物が存在する部屋の天井に配置されたマイクにより2人の対象人物A、Bの声と若干の雑音とを含んだ採取音が集音される。各マイクにより集音された採取音の時系列データが図1（A）に示す波形であるとする（なお、説明の都合上マイク数は7個としているが、本発明はこれに限定されるものでなくその数はもっと増やすこともできる）。

【0012】図1（A）に示すように、各マイクで集音される採取音の時系列データの波形における対象人物Aの声や対象人物Bの声に対応する部分は、各マイク毎に時間軸（横軸）に沿ってずれる。即ち、各対象人物と各マイクとの距離の大小により、対象人物の声がマイクに到達する時間が異なるためである。例えば、マイク1は対象人物Aに近く対象人物Bからは遠いため、該マイク1での時系列データは時間軸に沿って先に対象人物Aの声に対応する部分が表れ、後に対象人物Bの声に対応する部分が表れることになる。

【0013】抽出手段は、複数のマイクの各々により採

取された採取音の時系列データのうち複数の採取音の時系列データを選択する。ここでは、全てのマイクにより採取された採取音の時系列データを選択しても良いし、後述する請求項8記載の発明のように、オブジェクトの位置から所定距離以上離れたマイクにより採取された採取音の時系列データを除外し、それ以外の採取音の時系列データを選択しても良い。

【0014】そして、抽出手段は、選択した採取音の時系列データを、画像認識手段により認識されたオブジェクトの位置及び該選択した採取音を採取したマイクの位置に基づいて、オブジェクトが発する音が同期するようにシフトする。

【0015】例えば、図2の対象人物Aをオブジェクトとし、該対象人物Aの声の抽出を例にして説明すると、対象人物Aの頭部Pと各マイクとの距離を音速で除算することにより、対象人物Aの発した声に対する各マイクでの集音の遅延時間を求める。そして、図1(B)に示すように各マイクについて、該マイクにより集音された採取音の時系列データを時間軸に沿って遅延時間だけ遅延させた時系列データを求める。これにより、各マイクで対象人物Aの声に対応する部分が時間軸に沿ってほぼ同期する(同じ位相に揃う)ことになる。一方、当該対象人物Aの声以外の対象人物Bの声やその他の雑音に対応する部分は時間軸に沿って位相が揃わない状態のままである。

【0016】さらに、抽出手段は、シフトした採取音の時系列データを平均することにより、オブジェクトが発する音を抽出する。例えば、図1(B)に示すマイク1~7における時系列データを全て同期加算(重畳)し、加算した後の波形の振幅をマイク数「7」で除算する。これにより、相加平均した時系列データは図1(C)に示すように、対象人物Aの声以外の対象人物Bの声やその他の雑音に対応する部分ではその振幅が極めて小さくなりほぼ誤差の範囲内の振幅値となるため、対象人物Aの声に対応する部分のみが抽出されることになる。

【0017】以上のように、請求項1記載の発明によれば、オブジェクトの位置を認識し且つその位置に基づいてオブジェクトが発する音を周囲の雑音と弁別して抽出することができる。

【0018】また、第1の目的を達成するために、請求項2記載の発明は、請求項1記載の発明において、前記画像認識手段は、オブジェクトを含む領域の画像情報より該オブジェクトが発する方向も認識し、さらにオブジェクトの位置及び該オブジェクトが発する方向に基づいて、オブジェクトが発する音を良好に抽出可能な位置をオブジェクトの位置と改めて認識する、ことを特徴とする。

【0019】この請求項2記載の発明では、画像認識手段は、オブジェクトを含む領域の画像情報より、該オブジェクトが発する方向も認識する。例えば、図2に

示す対象人物Aが発する方向は、以下のようにして認識する。即ち、まず頭部Pを前述した要領で認識した後、その頭部Pの下に位置する胴体Sを認識し、その胴体Sにおいて胸幅L2は肩幅L1よりも小さいということから、対象人物Aは矢印V方向又はその逆方向を向いていると推定する。次に、頭部Pの表面における毛髪の占める比率は顔の位置する側よりも顔の位置しない側の方が高いという一般的特徴に基づいて、図2において紙面奥側の方が紙面前側よりも黒色の度合いが高いことから、頭部Pは矢印V方向を向いていると推定し、対象人物Aが発する方向は矢印V方向であると認識する。

【0020】さらに、画像認識手段は、オブジェクトの位置及び該オブジェクトが発する方向に基づいて、オブジェクトが発する音を良好に抽出可能な位置、即ち低周波域~高周波域にわたる全ての周波数成分をもれなく原音と略同一レベルで抽出できる位置(例えばオブジェクトの位置から前記音を発する方向に所定距離(数十cm)離間した位置)をオブジェクトの位置と改めて認識する。

【0021】このようにして認識されたオブジェクトの位置、即ちオブジェクトが発する音を良好に抽出可能な位置に基づいて、前述したような音の抽出を行うので、特にオブジェクトが発する音の指向性が強い場合には、より高い精度の音の抽出を行うことができる。

【0022】また、第1の目的を達成するために、請求項3記載の発明は、請求項1記載の発明において、オブジェクトが移動する場合、前記撮影手段はオブジェクトの移動に追従して該オブジェクトを含む領域を撮影する、ことを特徴とする。

【0023】この請求項3記載の発明ではオブジェクトが移動する場合、撮影手段はオブジェクトの移動に追従して該オブジェクトを含む領域を撮影する。これにより、画像認識手段は撮影された領域の画像情報より、移動するオブジェクトの位置を認識し、抽出手段は画像認識手段により認識されたオブジェクトの位置に基づいて前述した要領で当該移動するオブジェクトからの音を抽出する。従って、移動するオブジェクトに対しても該オブジェクトからの音を抽出することができる。

【0024】また、第1の目的を達成するために、請求項4記載の発明は、請求項1記載の発明において、オブジェクトが複数存在する場合、前記撮影手段は複数のオブジェクトを含む領域を撮影し、前記画像認識手段は撮影された領域の画像情報より複数のオブジェクトの各々の位置を認識し、前記抽出手段は該複数のオブジェクトの各々からの音を抽出する、ことを特徴とする。

【0025】この請求項4記載の発明では、オブジェクトが複数存在する場合、撮影手段は複数のオブジェクトを含む領域を撮影し、画像認識手段は撮影された領域の画像情報より複数のオブジェクトの各々の位置を個別に

認識する。そして、抽出手段は画像認識手段により認識された複数のオブジェクトの各々の位置に基づいて前述した要領で、複数のオブジェクトの各々からの音を抽出する。これにより、複数のオブジェクトに対しても、該複数のオブジェクトの各々からの音を抽出することができる。

【0026】また、第1の目的を達成するために、請求項5記載の発明は、請求項1記載の発明において、前記オブジェクト及び前記複数のマイクロフォンを含む領域において少なくとも音速及び音伝搬経路の一方に影響を及ぼす要因とされる音響環境状態を検出する検出手段を更に備え、前記抽出手段は、前記検出手段によって検出された音響環境状態が変化した場合に、変化した音響環境状態に基づいて、前記採取音の時系列データのシフトを補正することを特徴とする。

【0027】この請求項5記載の発明では、検出手段が、オブジェクト及び複数のマイクロフォンを含む領域において少なくとも音速及び音伝搬経路の一方に影響を及ぼす要因とされる音響環境状態、例えば温度や風力、風向を検出する。そして、抽出手段は、検出手段によって検出された音響環境状態が変化した場合に、変化した音響環境の状態に基づいて、例えば以下のようにして採取音の時系列データのシフトを補正する。

【0028】即ち、予め算出された検出温度に対応した音速と標準音速との比が、該検出温度に対応させて予め記憶された音速補正テーブルを参照することにより、検出温度に対応した音速と標準音速との比を得て、その比に基づいて各マイクでの集音の遅延時間を補正し、この補正された遅延時間に従って遅延操作を行う。又は、検出温度に対応した音速によって、オブジェクトの位置と各マイクの位置との距離を除算することにより、各マイクでの集音の遅延時間を補正し、この補正された遅延時間に従って遅延操作を行う。

【0029】また、例えば、検出した風力、風向の下でのオブジェクトの位置から各マイクの位置への音の伝搬経路を、予め各種の風力値、風向値を想定してシミュレーションすることにより求められた伝搬経路の曲がり

(変化)に関する情報に基づいて推定し、推定された伝搬経路に沿った距離を音速で除算することにより、各マイクでの集音の遅延時間を補正し、この補正された遅延時間に従って遅延操作を行う。

【0030】以上のようにして、音響環境状態の変化に応じて精度の高い音の抽出を行うことができる。

【0031】また、第1の目的を達成するために、請求項6記載の発明は、請求項1記載の発明において、前記抽出手段は、高音域の指向性に関する情報に基づき、高音域の採取音の時系列データを重み付けして平均することを特徴とする。

【0032】図11に音の周波数帯毎の成分が伝搬する領域を示す。高音域は殆ど音の方向(矢印D)にのみ伝

搬するのに対し、低音域ほどより広い範囲に伝搬することがわかる。即ち、音の指向性は、周波数帯により異なり、低音域ほど指向性が緩く、高音域は指向性が強くなるのが一般的である。よって、オブジェクトが音を発する方向に位置するマイクでは、低音域から高音域までほぼ全域の周波数成分が集音されるのに対し、その他のマイクでは低音域は集音されるものの高音域はあまり集音されない。

【0033】ところが、この請求項6記載の発明では、抽出手段は、高音域の指向性に関する情報に基づき、オブジェクトが音を発する方向に位置するマイクによる高音域の採取音と、他のマイクによる高音域の採取音と、の不均衡を是正すべく、重み付けして平均する。これにより、高音域が低音域よりも相対的に弱くなることを防止することができる。

【0034】また、第1の目的を達成するために、請求項7記載の発明は、請求項1記載の発明において、前記画像認識手段は、オブジェクトを含む領域の画像情報より該オブジェクトが音を発する方向、該オブジェクトの周辺に位置する音の反射面の位置及び向きもさらに認識し、前記抽出手段は、前記選択した採取音を採取したマイクロフォンの位置、オブジェクトの位置、該オブジェクトが音を発する方向、及び前記反射面の位置並びに向きに基づいて、前記選択した採取音の時系列データを、オブジェクトからの直接音又は前記反射面で反射された反射音の何れか一方が同期するようにシフトする、ことを特徴とする。

【0035】ところで、オブジェクトの発した音が所定の反射面に反射されてマイクにより集音された場合、それらの反射音は通常直接音に比べ非常に弱いため、加算平均することにより他の雑音成分と共に自動的に除去される。しかし、オブジェクトから近い位置であり且つ該オブジェクトが音を発する方向に、反射面が位置し、該反射面による反射音の進行方向にマイクロフォンが位置している場合等には、該反射面での反射音は直接音よりも大きくなるため、むしろ該反射面での反射音を集音した方がオブジェクトの発する音の抽出効果が高い。

【0036】そこで、この請求項7記載の発明では、画像認識手段は、オブジェクトを含む領域の画像情報より、該オブジェクトが音を発する方向、該オブジェクトの周辺に位置する音の反射面の位置及び向きもさらに認識する。抽出手段は、前記選択した採取音を採取したマイクロフォンの位置、オブジェクトの位置、該オブジェクトが音を発する方向、及び前記反射面の位置並びに向きに基づいて、前記選択した採取音の時系列データを、オブジェクトからの直接音又は前記反射面で反射された反射音の何れか一方が同期するようにシフトする。

【0037】例えば、反射面が、オブジェクトから近い位置であり且つ該オブジェクトが音を発する方向に位置しており、該反射面による反射音の進行方向にマイクロ

フォンが位置している場合には、反射面で反射された反射音の伝搬時間に応じた遅延操作を実行する。このように、オブジェクトの周辺の反射面等の配置状況に応じてより適切な音の抽出を行うことができる。

【0038】また、第1の目的を達成するために、請求項8記載の発明は、請求項1記載の発明において、前記抽出手段は、前記複数のマイクロフォンの各々により採取された採取音の時系列データのうち、前記オブジェクトの位置から所定距離以上離れて位置するマイクロフォンによって採取された採取音の時系列データを、選択の
10 対象から除外することを特徴とする。

【0039】一般的に音はその伝搬距離に応じて減衰するため、オブジェクトの発した音が長い伝搬距離を経てマイクに集音された場合に、当該マイクにより集音された採取音は、該オブジェクトの発した音の成分を少ししか含んでいないため、オブジェクトの音の時系列データを求める際に、該時系列データの形成に寄与する度合いが小さい。

【0040】そこで、この請求項8記載の発明では、抽出手段は、複数のマイクのうちオブジェクトの位置から
20 遠いマイク、即ち予め実験により求められた所定距離以上離れて位置するマイクによって採取された採取音の時系列データを、選択の対象から除外する。これにより、音の抽出の精度を低下させることなく、音の抽出に係る処理（抽出手段によるシフト及び平均処理）の負荷を軽減することができる。

【0041】なお、上記と同じ目的で、複数のマイクのうち、該マイクによって取られた採取音の音量が小さいマイク、即ち採取音の音量が予め実験により求められた所定音量レベルより小さいマイクによって取られた採取
30 音を、選択の対象から除外するようにしても良い。

【0042】また、第1の目的を達成するために、請求項9記載の発明は、請求項1記載の発明において、前記抽出手段により抽出されたオブジェクトが発する音を所定の音声認識装置へ出力する出力手段を更に備えたことを特徴とする。

【0043】ところで、請求項1記載の発明では、オブジェクトの近くに配置されたより多くのマイクで取られた採取音の時系列データに対して上記のようにシフト及び平均を行うことにより、信号対雑音比を改善させてオブジェクトの音を抽出することができる。しかも通常のマイクで集音した音よりも信号対雑音比が高い音を抽出することも可能である。このように良質な音は、音声認識装置への入力として活用することができる。
40

【0044】そこで、この請求項9記載の発明では、出力手段が抽出手段により抽出されたオブジェクトからの音を所定の音声認識装置へ出力する。これにより、音抽出装置により音の抽出が可能な領域内にいる人（一人又は複数人）が発した声を音声認識装置へ入力することができる。特に、身体の不自由な老人や障害者が音声認識
50

装置を用いて声により家庭内の電気製品等のスイッチのオン/オフ等を制御する場合に、適用することができる。

【0045】また、第2の目的を達成するために、請求項10記載の発明は、予め定められた位置に配置された広角固定焦点レンズを備え、音源としてのオブジェクトを含む領域を撮影する撮影手段と、前記撮影手段により撮影された領域の画像情報よりオブジェクトの位置を認識する画像認識手段と、予め定められた位置に配置され
10 前記オブジェクトが発する音を採取する複数のマイクロフォンと、前記複数のマイクロフォンの各々により採取された採取音の時系列データのうち複数の採取音の時系列データを選択し、選択した採取音の時系列データを、前記画像認識手段により認識されたオブジェクトの位置及び該選択した採取音を採取したマイクロフォンの位置に基づいて、オブジェクトが発する音が同期するようにシフトし、シフトした採取音の時系列データを平均することにより、オブジェクトが発する音を抽出する抽出手段と、を有することを特徴とする。

【0046】この請求項10記載の発明では、撮影手段は、予め定められた位置に配置された広角固定焦点レンズを備えている。これにより、オブジェクトが移動している場合、静止している場合を問わず、撮影手段（例えばテレビカメラ等）の向きをオブジェクトの移動に追従させて変更させることなく、オブジェクトを撮影することができる。また、物や人・動物等のオブジェクトは床や地面からの高さが概ね決まっており、更に、広角固定焦点レンズは焦点深度が大きいという特性を有しているので、撮影手段が焦点調整機構を有していなくても、即ち焦点調整を行うことなく、オブジェクトを撮影することが
20 できる。

【0047】このように、オブジェクトの移動に追従した撮影手段の向きの変更や焦点調整を行うことなく、オブジェクトの位置を速やかに認識することができる。また、上記のような撮影手段の向きの変更や焦点調整を行うための機械的な作動機構が不要となるので、撮影手段並びに音抽出装置の構造を簡単にできると共に、機械的な作動部を減らすことにより耐久性を向上させることができる。

【0048】なお、広角固定焦点レンズの配置位置としては、例えば、部屋の天井等の平面部の他、天井と壁との2面が作る角部や、天井と壁2面との計3面が作る角部とすることが
30 できる。

【0049】また、請求項10記載の発明では、上記のような撮影手段により撮影された領域の画像情報より、請求項1記載の発明と同様の要領で画像認識手段によってオブジェクトの位置を認識し、抽出手段によってオブジェクトが発する音を周囲の雑音と弁別して抽出することが
40 できる。

【0050】また、第2の目的を達成するために、請求

項11記載の発明は、請求項10記載の発明において、前記撮影手段は複数設けられており、各撮影手段は、前記広角固定焦点レンズによる結像点に配置されたエリアセンサを更に備え、前記画像認識手段は、該複数の撮影手段により撮影された各々異なる撮影情報を処理してオブジェクトの形状を認識する形状認識手段と、前記形状認識手段により認識されたオブジェクトの3次元座標を演算する3次元座標演算手段と、を含んで構成された、ことを特徴とする。

【0051】この請求項11記載の発明では、撮影手段は複数設けられており、各撮影手段は、広角固定焦点レンズによる結像点に配置されたエリアセンサを更に備えている。即ち、撮影手段によって、広角固定焦点レンズを介して撮影されたオブジェクトの像はエリアセンサ上に結像される。このようにして、複数の撮影手段により、オブジェクトを含む領域が各々異なる位置から撮影される。

【0052】形状認識手段により複数の撮影手段で撮影された各々異なる撮影情報が処理されてオブジェクトの形状が認識される。オブジェクトの形状を認識するには、例えば、後述する請求項12記載の発明のように、3次元空間をX軸、Y軸及びZ軸の各方向に沿って仮想的に細分割することにより得られる多数の立方体状の微小空間のうち、オブジェクトが占有する微小空間により形成される領域を求めることにより認識してもよいし、また、例えば、複数の撮影手段により各々異なる撮影情報を平面化した画像情報に変換し、該変換された画像情報から少なくともオブジェクトの正面、背面、左側面、右側面及び平面の画像情報を求め、該求められた画像情報を合成して認識してもよい。

【0053】3次元座標演算手段により形状認識手段で認識されたオブジェクト又は該オブジェクトの所定の部位の3次元座標が演算される。一般にオブジェクトは点とは限らず点の集合体であるので、オブジェクトが含む点の全ての3次元座標を全て演算してもよいし、オブジェクトと3次元空間とを画する境界に属する点の全ての3次元座標を演算してもよい。また、例えば特定の位置を予め設定（記憶）しておいて該特定の位置の3次元座標を演算することにより、高さや長さ等を求めても良い。

【0054】このようにして画像認識手段を構成する形状認識手段及び3次元座標演算手段によって、オブジェクトの3次元座標を速やかに求め、オブジェクトの位置を速やかに認識することができる。

【0055】また、第2の目的を達成するために、請求項12記載の発明は、請求項11記載の発明において、前記形状認識手段は、前記複数の撮影手段により撮影された各々異なる画像情報に基づいて、3次元空間をX軸、Y軸及びZ軸の各方向に沿って仮想的に細分割することにより得られる多数の立方体状の微小空間のうち、

オブジェクトが占有する微小空間により形成される領域を求めることによりオブジェクトの形状を認識することを特徴とする。

【0056】この請求項12記載の発明では、形状認識手段により、3次元空間をX軸、Y軸及びZ軸の各方向に沿って仮想的に細分割することにより得られる多数の立方体状の微小空間のうち、オブジェクトが占有する微小空間により形成される領域を求めることによりオブジェクトの形状を認識する。ここで、微小領域はエリアセンサの解像度の限界まで細分化することができる。このため、オブジェクトの形状を細部まで認識することができる。

【0057】また、第2の目的を達成するために、請求項13記載の発明は、請求項11記載の発明において、前記形状認識手段は、前記複数の撮影手段により撮影された各々異なる画像情報に基づいて、3次元空間をX軸、Y軸及びZ軸の各方向に沿って仮想的に細分割することにより得られる多数の立方体状の微小空間のうち、各撮影手段からオブジェクトを投影する視野角内に含まれる微小空間をそれぞれ抽出し、抽出した微小空間の全てに含まれる微小空間により形成される領域を求めることによりオブジェクトの形状を認識することを特徴とする。

【0058】上述した請求項11記載の複数の撮影手段により撮影された画像情報には、図19に示したように影（死角）の領域が生ずる。そこで、請求項13に記載したように、形状認識手段が、複数の撮影手段により撮影された各々異なる画像情報に基づいて、3次元空間をX軸、Y軸及びZ軸の各方向に沿って仮想的に細分割することにより得られる多数の立方体状の微小空間のうち、各撮影手段からオブジェクトを投影する視野角内に含まれる微小空間をそれぞれ抽出し、抽出した微小空間の全てに含まれる微小空間により形成される領域を求めることによりオブジェクトの形状を認識する。

【0059】このようにしてオブジェクトの形状を認識することにより、影（死角）の領域を排除することができ、オブジェクトの形状を正確に認識することができる。

【0060】また、第2の目的を達成するために、請求項14記載の発明は、請求項10記載の発明において、前記撮影手段は複数設けられており、各撮影手段は、前記広角固定焦点レンズによる結像点に配置されたエリアセンサを更に備え、前記画像認識手段は、各撮影手段のエリアセンサ上に結像された2次元座標を取得し、該取得した複数の2次元座標に基づいてオブジェクトの位置を認識することを特徴とする。

【0061】この請求項14記載の発明では、複数の撮影手段の各々のエリアセンサ上の点の2次元座標が分かれば3次元座標を逆算することができるので、画像認識手段が、各撮影手段のエリアセンサ上に結像された2次

10

20

30

40

50

元座標を取得し、該取得した複数の2次元座標に基づいてオブジェクトの位置（3次元座標）を正確に認識することができる。また、第2の目的を達成するために、請求項15記載の発明は、予め定められた位置に配置された広角固定焦点レンズ及び該レンズによる結像点に配設されたエリアセンサを備え、音源としてのオブジェクトを含む領域を撮影する撮影手段と、前記撮影手段の近傍に配置され、前記エリアセンサ上に結像するようにオブジェクトの像を反射する反射手段と、前記反射手段により反射されて前記エリアセンサ上に結像されたオブジェクト像、及び前記反射手段により反射されることなく前記エリアセンサ上に結像されたオブジェクト像の各々の該エリアセンサ上における2次元座標を取得し、該取得した複数の2次元座標に基づいてオブジェクトの3次元座標を演算することにより、オブジェクトの位置を認識する画像認識手段と、予め定められた位置に配置され前記オブジェクトが発する音を採取する複数のマイクロフォンと、前記複数のマイクロフォンの各々により採取された採取音の時系列データのうち複数の採取音の時系列データを選択し、選択した採取音の時系列データを、前記画像認識手段により認識されたオブジェクトの位置及び該選択した採取音を採取したマイクロフォンの位置に基づいて、オブジェクトが発する音が同期するようにシフトし、シフトした採取音の時系列データを平均することにより、オブジェクトが発する音を抽出する抽出手段と、を有することを特徴とする。

【0062】この請求項15記載の発明では、撮影手段が、予め定められた位置に配置された広角固定焦点レンズ及び該レンズによる結像点に配設されたエリアセンサを備えている。また、撮影手段により、オブジェクトを含む領域が撮影される。

【0063】反射手段は撮影手段の近傍に配置されている。この反射手段は、例えば図24の（G）～（L）に示したように、壁に沿って配置したり、L字型でもよく、また、湾曲したものであってもよい。反射手段によりエリアセンサ上に結像するようにオブジェクトの像が反射される。

【0064】そして、画像認識手段は、反射手段により反射されてエリアセンサ上に結像されたオブジェクト像、及び反射手段により反射されることなくエリアセンサ上に結像されたオブジェクト像の各々のエリアセンサ上における2次元座標を取得する。このようにして、撮影手段がたとえ1つであっても、その単一の撮影手段に備えたエリアセンサに複数のオブジェクト像が結像され、複数の2次元座標が取得されることになる。従って、上述した請求項14記載の発明と同様に、画像認識手段は、取得した複数の2次元座標に基づいてオブジェクトの位置（3次元座標）を正確に認識することができる。

【0065】このように反射手段により、もう1つのオ

ブジェクト像をエリアセンサ上に結像させることができるので、撮影手段が1つであってもオブジェクトの3次元座標を演算し、オブジェクトの位置を正確に認識することができる。

【0066】また、請求項15記載の発明では、上記のようにして画像認識手段によってオブジェクトの位置を認識し、請求項1記載の発明と同様の要領で抽出手段によってオブジェクトが発する音を周囲の雑音と弁別して抽出することができる。

10 【0067】また、第1及び第2の目的を達成するために、請求項16記載の発明は、請求項10乃至15の何れか1項に記載の発明において、前記画像認識手段は、オブジェクトを含む領域の画像情報より該オブジェクトが発する音の方向も認識し、さらにオブジェクトの位置及び該オブジェクトが発する音の方向に基づいて、オブジェクトが発する音を良好に抽出可能な位置をオブジェクトの位置と改めて認識することを特徴とする。

20 【0068】この請求項16記載の発明では、前述した請求項2記載の発明と同様に、オブジェクトの位置、即ちオブジェクトが発する音を良好に抽出可能な位置に基づいて、音の抽出を行うので、特にオブジェクトが発する音の指向性が強い場合やオブジェクトの音を発する部分（面）が大きい場合等には、より高い精度の音の抽出を行うことができる。

【0069】また、第1及び第2の目的を達成するために、請求項17記載の発明は、請求項10乃至15の何れか1項に記載の発明において、オブジェクトが複数存在する場合、前記撮影手段は複数のオブジェクトを含む領域を撮影し、前記画像認識手段は撮影された領域の画像情報より複数のオブジェクトの各々の位置を認識し、前記抽出手段は該複数のオブジェクトの各々からの音を抽出する、ことを特徴とする。

【0070】この請求項17記載の発明では、オブジェクトが複数存在する場合、前述した請求項4記載の発明と同様に、撮影手段は複数のオブジェクトを含む領域を撮影し、画像認識手段は撮影された領域の画像情報より複数のオブジェクトの各々の位置を認識する。そして、抽出手段は画像認識手段により認識された複数のオブジェクトの各々の位置に基づいて前述した請求項1記載の発明と同様の要領で、複数のオブジェクトの各々からの音を抽出する。これにより、複数のオブジェクトに対しても、該複数のオブジェクトの各々からの音を抽出することができる。

【0071】また、第1及び第2の目的を達成するために、請求項18記載の発明は、請求項10乃至15の何れか1項に記載の発明において、前記オブジェクト及び前記複数のマイクロフォンを含む領域において音速及び音伝搬経路の少なくとも一方に影響を及ぼす要因とされる音響環境状態を検出する検出手段を更に備え、前記抽出手段は、前記検出手段によって検出された音響環境状

態が変化した場合に、変化した音響環境状態に基づいて、前記採取音の時系列データのシフトを補正することを特徴とする。

【0072】この請求項18記載の発明では、抽出手段が、検出手段によって検出された音響環境状態が変化した場合に、変化した音響環境の状態に基づいて、前述した請求項5記載の発明と同様に採取音の時系列データのシフトを補正する。これにより、音響環境状態の変化に応じて精度の高い音の抽出を行うことができる。

【0073】また、第1及び第2の目的を達成するために、請求項19記載の発明は、請求項10乃至15の何れか1項に記載の発明において、前記抽出手段は、高音域の指向性に関する情報に基づき、高音域の採取音の時系列データを重み付けして平均することを特徴とする。

【0074】この請求項19記載の発明では、前述した請求項6記載の発明と同様に、抽出手段は、高音域の指向性に関する情報に基づき、オブジェクトが音を発する方向に位置するマイクによる高音域の採取音と、他のマイクによる高音域の採取音と、の不均衡を是正すべく、重み付けして平均する。これにより、高音域が低音域よりも相対的に弱くなることを防止することができる。

【0075】また、第1及び第2の目的を達成するために、請求項20記載の発明は、請求項10乃至15の何れか1項に記載の発明において、前記画像認識手段は、オブジェクトを含む領域の画像情報より該オブジェクトが音を発する方向、該オブジェクトの周辺に位置する音の反射面の位置及び向きもさらに認識し、前記抽出手段は、前記選択した採取音を採取したマイクロフォンの位置、オブジェクトの位置、該オブジェクトが音を発する方向、及び前記反射面の位置並びに向きに基づいて、前記選択した採取音の時系列データを、オブジェクトからの直接音又は前記反射面で反射された反射音の何れか一方が同期するようにシフトすることを特徴とする。

【0076】この請求項20記載の発明では、前述した請求項7記載の発明と同様に、例えば、反射面が、オブジェクトから近い位置であり且つ該オブジェクトが音を発する方向に位置しており、該反射面による反射音の進行方向にマイクロフォンが位置している場合には、反射面で反射された反射音の伝搬時間に応じた遅延操作を実行する。このように、オブジェクトの周辺の反射面等の配置状況に応じてより適切な音の抽出を行うことができる。

【0077】また、第1及び第2の目的を達成するために、請求項21記載の発明は、請求項10乃至15の何れか1項に記載の発明において、前記抽出手段は、前記複数のマイクロフォンの各々により採取された採取音の時系列データのうち、前記オブジェクトの位置から所定距離以上離れて位置するマイクロフォンによって採取された採取音の時系列データを、選択の対象から除外することを特徴とする。

【0078】この請求項21記載の発明では、前述した請求項8記載の発明と同様に、抽出手段は、複数のマイクのうちオブジェクトの位置から遠いマイク、即ち予め実験により求められた所定距離以上離れて位置するマイクによって取られた採取音を、選択の対象から除外する。これにより、音の抽出の精度を低下させることなく、音の抽出に係る処理（抽出手段によるシフト及び平均処理）の負荷を軽減することができる。

【0079】また、第1及び第2の目的を達成するために、請求項22記載の発明は、請求項10乃至15の何れか1項に記載の発明において、前記抽出手段により抽出されたオブジェクトが発する音を所定の音声認識装置へ出力する出力手段を更に備えたことを特徴とする。

【0080】この請求項22記載の発明では、前述した請求項9記載の発明と同様に、出力手段が抽出手段により抽出されたオブジェクトの音を所定の音声認識装置へ出力する。これにより、音抽出装置により音の抽出が可能な領域内にいる人（一人又は複数人）が発した声を音声認識装置へ入力することができる。

【0081】

【発明の実施の形態】

【第1の実施形態】以下、図面を参照して本発明の第1の実施形態を説明する。本第1の実施形態では、図2に示す所定の部屋50の中にいる対象人物Aの声のみを抽出する例を示す。

【0082】図2及び図3に示すように、第1の実施形態に係る音抽出装置10は、部屋50の天井52の所定位置に配置された複数のテレビカメラ16と、各テレビカメラ16に接続され該テレビカメラ16により撮影された画像情報に基づいて音の抽出位置を設定する抽出位置演算プロセッサ14と、天井52に略等間隔にマトリックス状に配置された複数（ n 個、図2には 8×8 個の例を示す）のマイク22を含むマイクロフォンアレイ部18と、各マイク22に接続され該マイク22により集音された音から対象人物の声を抽出する音声抽出ボード12と、抽出された音を出力するための出力端子ボード20と、を備えている。

【0083】各マイク22は、集音部24と、該集音部24に接続され音声信号のノイズカット及び増幅を行うアンプフィルタ26と、該アンプフィルタ26に接続されアナログ信号のデジタル信号への変換を行うA/Dコンバータ28と、を含んで構成されている。また、抽出位置演算プロセッサ14は、CPU14Aと、ROM14Bと、作業用記憶域として主に用いられるRAM14Cと、入出力コントローラ（以後、I/Oと称す）14Dと、を含んで構成されており、これらのCPU14A、ROM14B、RAM14C及びI/O14Dは互いにバス14Eによって接続されている。

【0084】また、音声抽出ボード12には、デジタル回線30を介して各マイク22と一対一の対応で接続さ

れ該マイク22から送信されてきた音声データを一時記憶するためのn個の入力バッファメモリi (i: 1, 2, ..., n) から構成された入力バッファメモリ群32と、各入力バッファメモリiに接続され音声抽出ボード12全体の制御等を行うプロセッサ34と、プロセッサ34に各々接続されプロセッサ34から出力された各マイク22に対応する音声データを一時記憶するためのn個の出力バッファメモリi (i: 1, 2, ..., n) から構成された出力バッファメモリ群44と、各出力バッファメモリiに接続され各出力バッファメモリiから出力された各マイク22に対応する音声データを加算するための加算器46と、加算器46に接続されデジタル信号のアナログ信号への変換を行うD/Aコンバータ48と、が設けられている。なお、プロセッサ34は、上記抽出位置演算プロセッサ14と同様に、CPU38、ROM40、RAM42及びI/O36を含んで構成されており、これらは互いにバス37によって接続されている。I/O36には上記の各入力バッファメモリi、各出力バッファメモリi及び抽出位置演算プロセッサ14が接続されている。また、プロセッサ34は、音抽出装置10における各構成機器の動作の同期を取るための制御信号等を各構成機器に送信するために、各構成機器、即ち各マイク22、入力バッファメモリ群32、出力バッファメモリ群44、加算器46、及びD/Aコンバータ48、の各々と制御信号線43を介して接続されている。なお、上記ROM40には、後述する音声抽出処理の制御プログラム、上記マイク22の各々の配置位置に関する位置情報、後述する遅延テーブル等が予め記憶されている。

【0085】また、出力端子ボード20は音声出力端子21を備えており、該音声出力端子21は音声抽出ボード12のD/Aコンバータ48に接続されている。

【0086】なお、抽出位置演算プロセッサ14に内蔵されたROM14Bには、上記テレビカメラ16の各々の配置位置を表す位置情報や後述する抽出位置演算処理の制御プログラム等が予め記憶されている。

【0087】次に、本第1の実施形態の作用を説明する。音抽出装置10の図示しないスタートボタンがオペレータによりオンされると、図4に示す抽出位置演算処理の制御ルーチンが抽出位置演算プロセッサ14のCPU14Aにより、図5に示す音声抽出処理の制御ルーチンが音声抽出ボード12のCPU38により、それぞれ実行される。なお、これらの制御ルーチンは何れも所定時間間隔で繰り返し実行される。

【0088】まず、図4に示す抽出位置演算処理の制御ルーチンから説明する。ステップ102では、各テレビカメラ16からの撮影情報を取り込む。次のステップ104では取り込んだ撮影情報より対象人物A(図2参照)の頭部Pの位置を演算する。なお、この時の位置としては、一例として図2に示すように部屋50を矢印X

方向、矢印Y方向、矢印Z方向の各々に沿って仮想的に等分割することにより得られる多数の直方体状の領域のうち何れの領域内に対象人物Aが位置するかを表す情報を用いることができる。図2では部屋50を各方向に16等分した場合を例として示している。即ちステップ104では、撮影画像から、表面の多くが毛髪で覆われており黒色部分が多い略球状である等のように人間の頭部に特有の特徴量を備えた領域を、対象人物Aの頭部Pに相当する領域として抽出し、抽出した領域の撮影画像上における位置に基づいて上述した仮想的な3次元座標上における頭部Pの位置を演算する。

【0089】また、ステップ104では対象人物Aの頭部Pの向きも推定を行う。即ち、まず図2に示す頭部Pの下に位置する胴体Sを認識し、その胴体Sにおいて胸幅L2は肩幅L1よりも小さいという一般的な特徴に基づいて、胸幅L2及び肩幅L1の大きさから対象人物Aが矢印V方向又はその逆方向を向いていると推定する。次に、頭部Pの表面における毛髪の占める比率は顔の位置する側よりも顔の位置しない側の方が高いという一般的特徴に基づいて、図2において紙面奥側の方が紙面手前側よりも黒色の度合いが高いことから、対象人物Aは矢印V方向を向いていると推定する。

【0090】次のステップ106では、ステップ104で求めた頭部Pの位置から矢印V方向に所定距離(例えば約30センチメートル)離間した位置を、対象人物Aに対する抽出位置として設定する。そして次のステップ108で、設定した抽出位置の位置情報を音声抽出ボード12へ送信する。

【0091】次に図5に示す、音声抽出ボード12に備えられたプロセッサ34のCPU38により実行される音声抽出処理の制御ルーチンを説明する。ステップ200では、上記のステップ108で抽出位置演算プロセッサ14から送信された抽出位置の情報を受信したか否かを判定する。抽出位置の情報を受信していない場合は制御ルーチンを終了し、抽出位置の情報を受信した場合はステップ202へ進む。ステップ202では、ROM40から取り出した各マイク22の設置位置情報と受信した抽出位置情報とに基づいて、抽出位置から所定距離以上離れた位置に設置されたマイク22を除外することにより抽出位置の音の抽出に適したマイク22を選択する。

【0092】一方、対象人物Aから発せられた音は、まずマイク22の集音部24により捉えられ、さらにアンプフィルタ26によりノイズがカットされると共に所定増幅率で増幅されて図1(A)に示すような音声信号となる。そしてそれらの音声信号はA/Dコンバータ28によりデジタル化された音声データに変換される。

【0093】そして、音声抽出処理のステップ203において、ステップ202で選択したマイク22の各々からデジタル回線30を介して、上記のようにして集音さ

れ変換された音声データを取り込み、該音声データを各マイク22に対応した入力バッファメモリiへ書き込む。即ち、図1(A)に示すような音声信号に対応する音声データが入力バッファメモリiへ書き込まれる。この際、入力バッファメモリiの所定の基準アドレスから順に書き込まれる。そして次に音声抽出処理ルーチンを実行する時には、前記基準アドレスから所定アドレスだけずらした新たな基準アドレスを設定し、該新たな基準アドレスから順に書き込まれる。そして入力バッファメモリiへの書き込みが3回終了したときには、次の4回目には新たな基準アドレスを入力バッファメモリiの先頭アドレスへ戻し、先頭アドレスから順に音声データが書き込まれる。このように入力バッファメモリiは所謂リングバッファとして使用される。

【0094】次のステップ212では、選択したマイク22のうちの1つのマイク22の位置と抽出位置との距離に対応する遅延時間を、ROM40に予め記憶された遅延テーブルから取り込む。なお、遅延テーブルは、部屋50の範囲内で変動しうる抽出位置の個々の抽出位置について、該抽出位置と各マイク22との距離を標準室温における音速で除算することにより予め求められた音の伝搬時間(遅延時間)を記録したテーブルであり、部屋50の範囲内で変動しうる抽出位置の候補数だけ予め準備されている。

【0095】次のステップ214では、当該1つのマイク22からの音声データを、前記所定の基準アドレス(即ち入力バッファメモリiへの書き込み開始アドレス)から遅延時間に対応するメモリアドレスだけずらしたアドレスを取り出しの先頭として、入力バッファメモリiから取り出す。これにより、対象人物Aが発した音が当該1つのマイク22に到達する前に入力バッファメモリiに書き込まれた音声データが切捨てられ、対象人物Aが発し当該1つのマイク22に到達した音を取り出されることになる。

【0096】そして次のステップ216では、取り出した音声データを、該1つのマイク22に対応する出力バッファメモリiへ書き込む。即ち、図1(B)に示すような音声信号に対応する音声データが出力バッファメモリiへ書き込まれる。なお、出力バッファメモリiも上記の入力バッファメモリiと同様に所謂リングバッファとして使用される。

【0097】そして、上記のステップ212、214、216は、選択したマイクの全てに対して実行される。ステップ212、214、216の処理が、選択したマイクの全てに対して実行されると、ステップ218で肯定され、ステップ220へ進み選択したマイクの各々に対応する音声データを加算器46により加算処理させる。

【0098】次のステップ222では、加算された音声データをINT(10g2M)の桁数だけ小数点位置を

上位にずらしてD/Aコンバータ48へ出力する。これにより、加算された音声データをマイク数Mで除算した場合と略同様の結果を得ることができる。なお、ここでは上記以外に、加算器46の演算結果をプロセッサ34で取り込み通常の除算を行っても良い。

【0099】その後加算器46から出力された音声データは、D/Aコンバータ48により図1(C)に示すようなアナログの音声信号に変換され、変換された音声信号は出力端子ボード20の音声出力端子21へ送出される。なお、この音声出力端子21には、音声再生装置等を接続することにより、抽出された対象人物Aの声を再生し聞くことができる。

【0100】以上の説明から明らかなように、複数の(図1の例では7個の)マイク22により集音された音に対し上記のような遅延操作及び加算平均を行うことにより、目的とする対象人物Aの声以外の雑音成分はその振幅が極めて小さくなってしまいうため、対象人物Aの声のみを抽出することができる。

【0101】また、上記の抽出位置演算処理(図4)及び音声抽出処理(図5)は、所定時間間隔で繰り返し実行される。これにより、対象人物Aが移動した場合、複数のテレビカメラ16により部屋50の内部が継続して撮影され、その画像情報に基づいて上記対象人物Aの移動に伴い変化する頭部Pの位置及び向きが求められ、その時点の頭部Pの位置及び向きに応じた抽出位置が設定される。そして音声抽出ボード12で、その抽出位置に応じて上記の音声抽出処理を行うことにより、対象人物Aが移動した場合でも、その声を抽出することができる。

【0102】なお、本第1の実施形態の音声抽出処理では、設定された抽出位置に近いマイク(例えば7個のマイク)を選択し、選択したマイクからの音声データのみを取り込み、入力バッファメモリへ書き込む例を示したが、一旦全ての(n個の)マイクからの音声データを取り込んで各入力バッファメモリへ書き込み、選択したマイク(例えば7個のマイク)からの音声データのみを、遅延時間に対応するメモリアドレスだけずらして入力バッファメモリから取り出すようにしても良い。

【0103】また、本発明の音声抽出処理では、抽出位置の近くに配置された多数のマイクで対象人物(又は対象物)の音を集音し、集音された音声信号を前述したように遅延操作及び加算平均を行うことにより、信号対雑音比の改善された音の抽出を行うことができる。しかも通常のマイクで集音した音よりも信号対雑音比が高い音を抽出することも可能である。このように良質な音は、音声認識装置への入力として活用することができる。即ち、音抽出装置で音の抽出が可能な領域内にいる人(一人又は複数人)が話した声を音声認識装置へ入力することができる。

【0104】〔第2の実施形態〕次に、本発明の第2の

実施形態を説明する。本第2の実施形態では、図2に示す所定の部屋50の中にいる対象人物Aの声と対象人物Bの声とを別々に抽出する例を示す。なお、第1の実施形態と同一の部分には同一の符号を付し、説明を省略する。

【0105】図6に示すように、本第2の実施形態における音抽出装置10には、第1の実施形態で説明した音声抽出ボード12が複数個(N個)備えられており、さらに各マイク22と各音声抽出ボード12とを接続するための音声データ中継ボード56が設置されている。また、抽出位置演算プロセッサ14は各音声抽出ボード12に備えられたプロセッサ34に接続されている。さらに、出力端子ボード20には各音声抽出ボード12に対応する音声出力端子21が設置されており、各音声出力端子21は対応する音声抽出ボード12のD/Aコンバータ48に接続されている。

【0106】次に本第2の実施形態の作用を説明する。音抽出装置10の図示しないスタートボタンがオペレータによりオンされると、図7に示す複数の抽出位置を対象とした抽出位置演算処理の制御ルーチンが抽出位置演算プロセッサ14のCPU14Aにより、図5に示す上記第1の実施形態と同じ音声抽出処理の制御ルーチンが2枚の音声抽出ボード12の各々のCPU38により、それぞれ実行される。

【0107】図7に示す抽出位置演算処理の制御ルーチンから説明する。なお、以下の説明では、対象人物A、Bをそれぞれ対象人物1、2と便宜上称する。ステップ102で各テレビカメラ16からの撮影情報を取り込み、次のステップ103では対象人物の数として「2」を変数Kに代入すると共に変数Lを「1」に初期設定する。

【0108】次のステップ105では、対象人物L(即ち対象人物1)の頭部の位置の演算と向きの推定を上記第1の実施形態と同じ要領で行い、次のステップ107では当該対象人物Lの声を抽出するための抽出位置L(即ち抽出位置1)を設定する。そして、次のステップ109でその抽出位置Lの情報を、対応する音声抽出ボードLへ送信する。

【0109】次のステップ110では変数Lが対象人物の数を示す変数Kに等しいか否かを判定することにより、全ての対象人物に対して上記ステップ105、107、109の処理が完了したか否かを判定する。この場合初めは否定され、ステップ112へ進み、変数Lを増分1だけインクリメントする。これにより変数Lの値は「2」となる。

【0110】その後ステップ105へ戻り対象人物L(即ち対象人物2)に対して、上記のステップ105、107、109の処理を行う。それらの処理が完了すると、ステップ110では、変数Lと変数Kとが等しいため肯定され、制御ルーチンを終了する。

【0111】対象人物1、2にそれぞれ対応する音声抽出ボード12では、上記ステップ109により抽出位置演算プロセッサ14から送信された抽出位置1又は抽出位置2の情報をそれぞれ受信し、該受信した情報に基づいて、第1の実施形態と同じ図5に示す音声抽出処理を実行する。説明は省略するが、対象人物1、2にそれぞれ対応する各音声抽出ボード12での音声抽出処理により、対象人物1、2の声をそれぞれ独立に抽出することができる。

10 【0112】なお、本第2の実施形態では、複数の音声抽出ボード12を備え、各音声抽出ボード12で1つの抽出位置から音を抽出する例を示したが、音声抽出の即時性があまり高く要求されない場合等には、単一の音声抽出ボード12において複数の抽出位置の各々に対する音声抽出処理を順次実行するようにしても良い。

20 【0113】〔第3の実施形態〕次に、本発明の第3の実施形態を説明する。本第3の実施形態では、図2に示す部屋50内の温度変化による影響を考慮して該部屋50の中にいる対象人物Aの声のみを抽出する例を示す。なお、第1の実施形態と同一の部分には同一の符号を付し、説明を省略する。

30 【0114】図8に示すように、本第3の実施形態における音抽出装置10は、複数の温度センサ58を備えており、該温度センサ58は部屋50内の複数の温度測定地点に1台ずつ設置されている。各温度センサ58はプロセッサ34内のI/O36に接続されている。また、プロセッサ34内のROM40には、温度センサ58により測定された部屋50の複数の温度測定地点の温度に基づいて、部屋50内の温度分布を推定するための温度分布情報が予め記憶されている。

【0115】次に本第3の実施形態の作用を説明する。音抽出装置10の図示しないスタートボタンがオペレータによりオンされると、図4に示す上記第1の実施形態と同じ抽出位置演算処理の制御ルーチンが抽出位置演算プロセッサ14のCPU14Aにより、図9に示す音声抽出処理の制御ルーチンが音声抽出ボード12のCPU38により、それぞれ実行される。以下では、抽出位置演算処理の説明は省略し、図9を用いて本第3の実施形態における音声抽出処理について説明する。

40 【0116】ステップ203では、選択したマイク22の各々について、該マイク22からの音声データの取り込みと、取り込んだ音声データの入力バッファメモリiへの書き込みと、を実行し、次のステップ204では選択したマイク22のうちの1つのマイク22に対して、該マイク22と抽出位置との距離を算出する。

50 【0117】次のステップ205で複数の温度センサ58の各々から部屋50の所定の温度測定地点における温度を取り込み、次のステップ206では、取り込んだ複数の温度測定地点の温度に基づき、ROM40に記憶された上記の温度分布情報を参照することによって、部屋

50内の温度分布を推定し、抽出位置から発せられた音が該マイク22に到達するまでの音伝搬経路上の平均温度を算出する。

【0118】次のステップ207で音伝搬経路上の平均温度に基づいて該音伝搬経路上の音速を算出し、次のステップ208では、ステップ204で算出した該マイク22と抽出位置との距離を、ステップ207で算出した音速で除算することにより、該マイク22に到達する音の伝搬時間、即ち該マイク22についての遅延時間を算出する。そして次のステップ209では、算出した遅延時間を該マイク22の識別番号と対応付けた形で、RAM42に確保された遅延テーブルへ記憶する。なお、本第3の実施形態における遅延テーブルは、算出した各マイク22についての遅延時間を一時記憶するための一時記憶域として使用される。

【0119】上記ステップ204～209は、選択したマイク22の各々に対して実行される。選択したマイク22の全てに対して実行が完了すると、選択したマイク22の各々についての遅延時間が記録された遅延テーブルが完成する。その後、上記第1の実施形態と同様に、ステップ214で1つのマイク22からの音声データを上記遅延テーブルから得られた該マイク22に対する遅延時間に対応するメモリアドレスだけずらして入力バッファメモリiから取り出す。次のステップ216では取り出した音声データを出力バッファメモリiへ書き込む。

【0120】これらのステップ214、216の処理を、選択したマイク22の全てに対して実行完了すると、ステップ218で肯定され、ステップ220へ進む。ステップ220、222では、選択したマイク22の各々における音声データを加算平均し、D/Aコンバータ48へ出力する。音声データは、D/Aコンバータ48によりアナログの音声信号に変換され、変換された音声信号は出力端子ボード20の音声出力端子21へ出力される。

【0121】このように本第3の実施形態によれば、部屋50内の温度の変化に応じて精度の高い音の抽出を行うことができる。

【0122】なお、本発明の音抽出装置10は、上記と同様の要領で風（風向、風力）の影響による音の伝搬経路の曲がりを考慮して音の抽出を行うことができる。例えば図10に示すように、矢印R方向に走行する列車64が鉄橋66を渡る時に鉄橋66の特定の測定部位66Aが発するきしみ音を抽出する場合について説明する。この場合、屋外の音響環境であるので、音の伝搬に対し温度以外に風の影響が生じる。例えば、鉄橋66の測定部位66Aから発せられ1つのマイク22Aに到達する音の伝搬経路が、破線K1で示す直線経路ではなく実線K2で示す曲線経路となり、音の伝搬経路長L1（曲線経路の長さ）が測定部位66Aとマイク22Aとの距離

L2（直線経路の長さ）よりも長くなる。そこで、音抽出装置10では、風力計60により風力を、風向計62により風向を、それぞれ検出する。そして、音の伝搬経路が風力及び風向の影響でどのような経路（曲線経路）に変わり、その伝搬経路長L1は距離L2に比べどの程度長くなるかを、抽出位置演算プロセッサ14又は音声抽出ボード12のプロセッサ34で演算して求め、求められた伝搬経路長L1に基づいてマイク22Aにおける遅延時間を算出する。他のマイク22についても同様に音の伝搬経路長を求め、遅延時間を算出する。そして、算出した遅延時間に基づいて、以後の遅延操作及び加算平均を行い、測定部位66Aから発せられる音の抽出を行う。このようにして、風（風向、風力）の影響による音の伝搬経路の曲がりを考慮して音の抽出を行うことができる。

【0123】〔第4の実施形態〕次に、本発明の第4の実施形態を説明する。本第4の実施形態では、音声における周波数による指向性の違いを考慮した上で図11に示す部屋50の中にいる対象人物Cの声を抽出する例を示す。なお、第1の実施形態と同一の部分には同一の符号を付し、説明を省略する。

【0124】本第4の実施形態における音抽出装置10の構成は、前述した第1の実施形態における音抽出装置10の構成と同一であるので、その説明は省略する。但し、音声抽出ボード12のプロセッサ34内のROM40には、後述する重み付け定数が記録された重み付けテーブルが予め記憶されている。

【0125】次に本第4の実施形態の作用を説明する。まず、音における周波数による指向性の違いについて説明する。図11に示すように音の指向性は、周波数により異なり、低い周波数ほど指向性が緩く、高い周波数では指向性が強くなるのが一般的である。従って、対象人物Cが声を発する方向Dに位置するマイクでは、低周波数から高周波数までほぼ全域の周波数の音が集音されるのに対し、その他のマイクでは低周波数の音は集音されるものの高周波数の音はあまり集音されないことになる。

【0126】そこで、本第4の実施形態では、上記の方向Dに位置するマイク22により集音された採取音の高音域の音量と、他のマイク22により集音された採取音の高音域の音量と、の不均衡を是正すべく、両者に重み付け操作をすることにより、上記の課題を解消する例を示す。

【0127】なお、抽出位置演算処理は第1の実施形態と同一であるので説明は省略し、音声抽出処理について図12を用いて説明する。

【0128】ステップ200、202、203では、第1の実施形態と同様に抽出位置演算プロセッサ14から受信した抽出位置情報に基づいてマイクを選択し、選択したマイクからの音声データの取り込みと該音声データ

の入力バッファメモリ i への書き込みとを行う。次のステップ 213 では、1つのマイク 22 に対する抽出位置の相対的位置に対応した遅延テーブルから遅延時間を取り込むと共に、該マイク 22 に対する抽出位置の相対的位置及び音の方向に対応した重み付けテーブルから重み付け定数を取り込む。なお、人物 C が声を発する方向 D に位置するマイク 22 に対応する重み付け定数には、方向 D から外れた位置に位置するマイク 22 に対応する重み付け定数よりも相対的に小さい値が設定されている。

【0129】次のステップ 214 では第 1 の実施形態と同様に、該マイク 22 からの音声データを上記遅延時間に対応するメモリアドレスだけずらして入力バッファメモリ i から取り出し、次のステップ 217 では取り出した音声データにおける高周波数成分に対して、上記重み付け定数に従って重み付けして（レベルの増幅又は減縮を行って）出力バッファメモリ i へ書き込む。

【0130】上記ステップ 213、214、217 は、選択したマイク 22 の各々に対して実行される。こうして、方向 D に位置するマイク 22 により集音された採取音の高周波数成分はレベルが減縮される一方、方向 D から外れた位置に位置するマイク 22 により集音された採取音の高周波数成分はレベルが増幅される。

【0131】次のステップ 220、222 では選択したマイク 22 の各々における音声データを加算平均し、D/A コンバータ 48 へ出力する。音声データは、D/A コンバータ 48 によりアナログの音声信号に変換され、変換された音声信号は出力端子ボード 20 の音声出力端子 21 へ出力される。

【0132】本第 4 の実施形態によれば、方向 D に位置するマイクにより集音された採取音の高周波数成分と、方向 D から外れた位置に位置するマイク 22 により集音された採取音の高周波数成分と、でレベルの不均衡が改善され、高周波数の音の有する強い指向性に起因した高周波数の音の低周波数の音に対する相対的なレベル低下を防止することができる。

【0133】なお、上記第 1 乃至第 4 の実施形態では、対象人物（又は対象物）から発せられマイクに直接到達する直接音のみを抽出する例を示した。一般に、反射面としての壁面等から反射された後マイクに到達した反射音は、その大きさが上記直接音に比べ非常に小さいため、加算平均を行うことにより他の雑音成分と共に除去される。

【0134】しかし、壁面が対象人物から近い位置であり且つ該対象人物が音を発する方向に位置している場合には、壁面での反射音は直接音よりも大きくなるため、むしろ該反射音を集音した方が対象人物が発する音の抽出効果が高いと言える。

【0135】従って、テレビカメラ 16 により撮影された画像情報に基づいて、壁面が対象人物から近い位置であり且つ該対象人物が音を発する方向に位置していると

認識された場合、プロセッサ 34 の CPU 38 で実行される音声抽出処理において、各マイクについての遅延時間を演算するためのマイクと抽出位置との距離として両者の直接の距離ではなく、壁面で反射された反射音の伝搬距離を採用し、該反射音の伝搬距離に応じた遅延時間を算出し、そしてこの反射音の伝搬距離に応じた遅延時間に従って遅延操作を行えば良い。

【0136】これにより、対象人物から各マイクに直接到達した音は雑音成分として除去され、代わりに壁面で反射された後各マイクに到達した反射音が対象人物の音として抽出される。このように、各マイクに到達した反射音の方が直接音よりも対象人物（対象物）の音の抽出のために適当である場合には、反射音の方を抽出することができる。

【0137】なお、本発明の音抽出装置は、上記の各種の実施形態以外にも以下のように適用することができる。例えば、講演会場において聴衆の中の質問者の声を拡声する場合、聴衆を複数のテレビカメラで撮影しておき、係員が該質問者の映った画面で該質問者の口元付近をマウス等でポイントすると、抽出位置演算プロセッサが該質問者の口元付近を抽出位置に設定する。そして、音声抽出ボードでその抽出位置からの音の抽出を行い、所定のスピーカから抽出された音を出力する。これにより、聴衆の中の質問者の位置まで都度マイクを持っていく必要は無くなり、スムーズな講演会の進行の一助となる。

【0138】また、例えば図 10 に示す列車 64 のように移動経路の定まった移動体から発せられる音を時間の経過と共に抽出し続ける（トレースする）場合には、移動経路上に略等間隔の複数の抽出位置（例えば鉄橋の部位 66B、66C、66D）を予め設定し、これらの抽出位置における音を時間の経過に沿って順に抽出すれば良い。こうすれば、抽出位置を設定するために移動体の動きをテレビカメラ 16 により撮影された画像から把握する処理は不要となり、移動体の速い動きに追従して音をトレースすることができる。

【0139】〔第 5 の実施形態〕次に、本発明に係る第 5 の実施形態を説明する。本第 5 の実施形態では、オブジェクトの音の抽出を行うにあたり、広角固定焦点レンズを備えた複数のテレビカメラによりオブジェクトを含む画像を撮影し、その画像データに基づいてオブジェクトの位置を認識する例を示す。

【0140】図 13 に示すように、天井 52 には、テレビカメラ 16 が複数台（一例として 4 台）設置されており、各テレビカメラ 16 には、広角固定焦点レンズとしての魚眼レンズ 16A が設置されている。各魚眼レンズ 16A の視野角は 90° 以上に予め設定されている。よって、オブジェクトが移動している場合、静止している場合を問わず、テレビカメラ 16 を動かさずに該オブジェクトを撮影可能となっている。

【0141】この魚眼レンズとしては、例えば、等距離射影 (equidistance projection) 型、立体射影型、等立体角射影型、正射影型等の種々の種類があり、本実施形態では何れの魚眼レンズも使用可能であるが、以下、等距離射影型の魚眼レンズを使用した例を説明する。また、各テレビカメラ16はCCD (Charge-Coupled Device) エリアイメージセンサ16B (図18参照) を備えている。また、物や人・動物等のオブジェクトは床や地面からの高さが概ね決まっており、更に、広角固定焦点レンズとしての魚眼レンズ16Aは焦点深度が大きいという特性を有しているため、テレビカメラ16が焦点調整機構を有していなくても、鮮明にCCDエリアイメージセンサ16B上にオブジェクト像を結像することができ、このようにして、複数のテレビカメラ16の各々により、オブジェクトを含む所定の領域が各々異なる位置から撮影される。

【0142】次に、本第5の実施形態の作用を説明する。オペレータが、オブジェクトとして対象人物Aを指定し、音声抽出装置10の図示しないスタートボタンをオンすると、前述した第1の実施形態と同じ図5に示す音声抽出処理の制御ルーチンが音声抽出ボード12のCPU38により実行開始されると共に、図14に示す抽出位置演算処理の制御ルーチンがCPU14Aにより実行開始される。以下では、音声抽出処理の説明は省略し、図14～24を用いて本第5の実施形態における抽出位置演算処理について説明する。

【0143】図14に示すステップ120では、オブジェクト分別処理が行われる。このオブジェクト分別処理では、図15に示すサブルーチンが実行される。図15のステップ140では、オブジェクト (対象人物A) が部屋50内に存在しないときの画像データAをROM14Bから読み出し、次のステップ142において各テレビカメラ16で撮影した画像データBを各々取り込んでRAM14Cに記憶する。次のステップ144では画像データBと画像データAとの差をとって部屋50に存在する対象人物Aを認識する (図17参照)。

【0144】次にステップ146では所定時間Tのタイ

$$\theta = \tan^{-1} (R/H)$$

$$\theta' = \tan^{-1} \{R/(H-h)\}$$

$$r = f \theta$$

$$r' = f \theta'$$

従って、高さh及び距離Rは次式 (5) 及び (6) により求めることができる。

$$h = H \{1 - \tan(r/f) / \tan(r'/f)\} \dots (5)$$

$$R = H \tan(r/f) \dots (6)$$

なお、距離H及び焦点距離fは予め定められており、式 (5) 及び式 (6) はROM14Bに記憶されている。このため、このステップ122では、式 (5) をROM14Bから読み出して1台のテレビカメラ16のCCD

*マを設定し、次のステップ148では、所定時間Tの時間待ちに入り、タイムアウトになると、ステップ150へ進む。

【0145】ステップ150では、各テレビカメラ16で撮影した画像データC (即ち、画像データBから所定時間T経過後の画像データ) を取り込む。そして、次のステップ152では、RAM14Cに記憶した画像データBを読み出し、画像データBと画像データCとを比較し、次のステップ154ではこの比較結果に基づいて対象人物Aが移動しているか否かを判断する。

【0146】対象人物Aが移動していない (静止している) 場合は、ステップ154で否定判断され図14のメインルーチンへリターンする。一方、対象人物Aが移動している場合は、ステップ154で肯定判断されステップ156へ進み、画像データBと画像データCとの差異から対象人物Aの進行方向を求め (図17参照)、求めた進行方向から対象人物Aの前後を判断する。そして、次のステップ158で、対象人物Aの進行方向及び前後に関する情報をRAM14Cに記憶し、図14のメインルーチンへリターンする。

【0147】次のステップ122では、対象人物Aの位置及び高さを演算する。図18に示すように、点Oに固定された等距離射影型魚眼レンズ16Aの焦点距離をf、点Oから部屋50の床面54に垂直に下ろした点Qまでの距離をH、点Qから対象人物Aの床面54上の点Pまでの距離をR、対象人物Aの高さ (対象人物Aの天井方向の先端を点P' としたときに点P' と点Pとの距離) をhとする。また、点POQのなす角を θ 、点P'OQのなす角を θ' 、CCDエリアイメージセンサ16BのCCD面上のオブジェクト像の高さに対応する距離をh'、オブジェクト像h'のうち点P'に対応して結像した点をp、オブジェクト像h'のうち点P'に対応して結像した点をp'、CCD面の画像中心 (CCD面の中心) oから点pまでの距離をr、CCD面の画像中心oから点p'までの距離をr' とすると、角 θ 、 θ' 、距離r、r' は次式 (1) ~ (4) により求めることができる。

【0148】

$$\dots (1)$$

$$\dots (2)$$

$$\dots (3)$$

$$\dots (4)$$

※【0149】

面上の情報から高さhを演算し、式 (6) を読み出して2台のテレビカメラ16のCCD面上の情報からそれぞれ距離Rを求め、求めた2つの距離Rから対象人物Aの2次元位置を演算する。

【0150】次のステップ124では、上記ステップ122で演算した位置を中心として3次元空間をX方向、Y方向及びZ方向に沿って仮想的に細分割したマトリクス状の微小空間（以後、ボクセルと称す）を設定する。これにより、画像データCはボクセルの集合体に変換される。図19は4台のテレビカメラA、B、C及びDから対象人物Aを投影したとき、対象人物Aが占有するボクセルを概念的に示す。

【0151】即ち、各テレビカメラから対象人物Aを投影したときの、対象人物Aの視野角内に位置するボクセルは、影（死角）の部分 R_A 、 R_B 、 R_C 、 R_D も含めて、対象人物Aが占有するボクセルとして設定される。なお、ボクセルは、CCDエリアイメージセンサ16Bの解像度の限界まで細分割することが可能である。

【0152】次のステップ126では、画像データのうち対象人物Aが占有するボクセルを、対象人物Aの高さ h に基づいて以下のようにして限定する第1次絞り込みを行う。

【0153】対象人物Aの高さ h は、成人の平均身長から概ねその値が予め設定可能であるので、図20(A)～(D)に示すように各テレビカメラから対象人物Aを投影したときに対象人物Aの視野角内に位置するボクセルのうち、高さが0～ h の範囲のものを、対象人物Aが占有するボクセルとして絞り込む。なお、ここで絞り込まれたボクセルで形成される領域を第1次絞り込み領域とする。

【0154】次にステップ128では各画像データにおける第1次絞り込み領域から、それら全てに重複する領域に絞り込む第2次絞り込みを行う。これにより、図19に示した影の領域 R_A 、 R_B 、 R_C 及び R_D は対象人物Aが占有するボクセルから排除され、図21に示すように、対象人物Aが占めるボクセル70に絞りこまれる。次のステップ130では、このボクセル70によって、オブジェクトの位置及び形状を正確に認識する。なお、ボクセルはCCDエリアイメージセンサ16Bの解像度の限界まで細分化することができるので、オブジェクトの形状を細部まで認識することも可能である。

【0155】次のステップ132では、図22に示すように、ボクセル70の高さ、太さ等の寸法や、予めROM14Bに記憶された頭部の色差、目、鼻、口、耳の位置、腕の長さや位置、つま先の向き、関節の自由度等の*

$$\begin{aligned}\alpha_1 &= f \theta_1 \cos(\pi - \phi) = -f \theta_1 \cos \phi \\ \beta_1 &= f \theta_1 \sin(\pi - \phi) = f \theta_1 \sin \phi \quad \dots (7)\end{aligned}$$

である。ここで、

$$\begin{aligned}\sin \phi &= y / \{ (x - X)^2 + y^2 \}^{1/2} \quad \dots (8) \\ \cos \phi &= (x - X) / \{ (x - X)^2 + y^2 \}^{1/2}\end{aligned}$$

であるので、 α_1 、 β_1 は、

$$\alpha_1 = -f \theta_1 (x - X) / \{ (x - X)^2 + y^2 \}^{1/2} \quad \dots (9)$$

$$\beta_1 = f \theta_1 y / \{ (x - X)^2 + y^2 \}^{1/2} \quad \dots (10)$$

として求めることができる。式(10)を式(9)で割

*人間の特征に関する情報、さらに対象人物Aが移動している場合にはRAM14Cに記憶した対象人物Aの進行方向や前後に関する情報に基づいて、ダミーモデル72へ変換する。

【0156】次のステップ134では、図16に示す抽出位置設定処理のサブルーチンを実行する。図16のステップ160では、対象人物Aの頭部を撮影対象とした所定数（一例として2台）のテレビカメラの選択を行い、選択した各テレビカメラのCCD面上における対象人物Aの頭部の位置に対応する2次元座標を取り込む。なお、テレビカメラの選択においては、例えば、対象人物Aを写した際のオブジェクト像が大きい順に選択してもよいし、対象人物Aの正面を捉えたテレビカメラを選択してもよい。また、選択された2台のテレビカメラをそれぞれカメラL、カメラRとする。

【0157】次のステップ162では3次元座標を演算する。図23に示すように、カメラLの3次元座標 C を $(X, 0, Z)$ 、カメラRの3次元座標 C' を $(X', 0, Z)$ とする。また、対象人物Aの頭部の位置に対応するカメラLのCCD面上の座標 P_L を (α_1, β_1) 、カメラLのCCD面の画像中心 O_L から座標 P_L までの距離を r 、対象人物Aの頭部の位置に対応するカメラRのCCD面上の座標 P_R を (α_1', β_1') 、カメラRのCCD面の画像中心 O_R から座標 P_R までの距離を r' 、座標 P_L 及び座標 P_R から出た光を仮想したときに2つの光が交わる点、即ち、対象人物Aの頭部の3次元座標 P を (x, y, z) とする。

【0158】また、カメラLの3次元座標位置からZ軸に平行に下ろした垂線の足と、点Pを含みZ軸に垂直な平面と、の交点Sの座標を $(X, 0, z)$ とし、カメラRの3次元座標位置からZ軸に平行に下ろした垂線の足と、点Pを含みZ軸に垂直な平面と、の交点S'の座標を $(X', 0, z)$ とする。更に、点PCSのなす角を θ_1 、点PC'S'のなす角を θ_1' 、点PSS'のなす角を ϕ 、点PS'S'のなす角を ϕ' とする。

【0159】CCD面上で画像中心 O_L から像までの距離 r は、上述した式(3)により、

$$r = f \theta_1$$

として求められる。

【0160】また、それぞれ α_1 、 β_1 は、

ることにより、

33

$$y = (\beta_1 / \alpha_1) (X - x)$$

同様に、

$$y = (\beta_1' / \alpha_1') (X' - x) \quad \dots (12)$$

式(11)と式(12)とからyを消去して、

$$x = (\alpha_1 \beta_1' X' - \alpha_1' \beta_1 X) / (\alpha_1 \beta_1' - \alpha_1' \beta_1) \quad \dots (13)$$

により3次元座標PのX座標を求めることができる。

*消去して、

【0161】次に、式(11)と式(13)とからxを*

$$y = \beta_1 \beta_1' (X - X') / (\alpha_1 \beta_1' - \alpha_1' \beta_1) \quad \dots (14)$$

により3次元座標PのY座標を求めることができる。 ※10※ 【0162】ところで、

$$\theta_1 = \tan^{-1} [\{ (x - X)^2 + y^2 \}^{1/2} / (Z - z)]$$

であるので、式(7)、式(8)から

$$\beta_1 / (f \sin \phi) = \tan^{-1} [\{ (x - X)^2 + y^2 \}^{1/2} / (Z - z)]$$

従って、

$$z = Z - [\{ (x - X)^2 + y^2 \}^{1/2} / \tan [(\beta_1 / f) \times \{ (x - X)^2 + y^2 \}^{1/2} / y]] \quad \dots (15)$$

また、式(11)から

$$\{ (x - X)^2 + y^2 \}^{1/2} = (x - X) \times \{ 1 + (\beta_1 / \alpha_1)^2 \}^{1/2}$$

式(11)と式(14)とから

$$(x - X) = (X' - X) / \{ 1 - (\alpha_1' / \alpha_1) \times (\beta_1 / \beta_1') \}$$

であるので、式(15)は、

$$z = Z + [(X' - X) \times \{ 1 + (\beta_1 / \alpha_1)^2 \}^{1/2} / \{ 1 - (\alpha_1' / \alpha_1) \times (\beta_1 / \beta_1') \}] / \tan \{ (\alpha_1^2 + \beta_1^2)^{1/2} / f \} \quad \dots (16)$$

と表すことができ、3次元座標PのZ座標を求めることができる。

【0163】なお、各テレビカメラ16の3次元座標は予め定められているので、ステップ162では、ROM 14Bから式(13)、(14)及び(16)を読み出し、ステップ160で取り込んだカメラLのCCD面上の座標PL (α_1 , β_1)及びカメラRのCCD面上の座標PR (α_1' , β_1')の値を、式(13)、(14)及び(16)に代入することにより、対象人物Aの頭部の3次元座標P(x, y, z)を求めることができる。

【0164】次のステップ164では、前述した第1の実施形態と同様に(図4のステップ104における処理と同様に)対象人物Aの頭部の向きを推定する。次のステップ166では、ステップ162で求めた頭部の位置から矢印V方向(図13参照)に所定距離(例えば約30センチメートル)離間した位置を、対象人物Aに対する抽出位置として設定する。そして次のステップ168で、前記設定した抽出位置の位置情報を音声抽出ボード12へ送信してリターンする。

【0165】このように本第5の実施形態によれば、広角固定焦点レンズ16Aを用いて撮影しているので、テレビカメラ16を動かしたり焦点調整を行う必要はない。このため、オブジェクト(対象人物A)を捉えるま

での時間を短縮することができ、速やかにオブジェクトの位置を認識することができる。

【0166】また、テレビカメラの向きを変えたり焦点を調整調整するための機構が不要となるので、オブジェクトを捉える作業を自動化することができると共に、駆動部分がなくなるので、テレビカメラの耐久性や信頼性を高めることができる。

【0167】また、1つのオブジェクトに対して複数のテレビカメラにより撮影しているので、例えば家具等の視野を遮る障害物や他のオブジェクトが存在していても、3次元座標を演算することができる。

【0168】また、テレビカメラは3次元空間を構成する部屋の天井に配置されているので、壁面を有効に使用することができる。

【0169】なお、本第5の実施形態では複数のテレビカメラ16を天井52に配置したが、図24の(A)～(F)に示すように、壁の近くに配置したり壁に埋め込んで配置してもよく、天井と壁で構成される2面のコーナー部や天井と2面の壁で構成される3面のコーナー部に配置してもよい。更に、図24の(M)～(O)に示すように、等距離射影型魚眼レンズ16Aを部屋の中心に向けるようにしてもよい。

【0170】また、本第5の実施形態では、図15に示すオブジェクト分別処理におけるステップ140で、オ

34

$$\dots (11)$$

$$\dots (12)$$

$$\dots (13)$$

$$\dots (14)$$

$$\dots (15)$$

$$\dots (16)$$

30

40

50

ブジェクトが部屋50内に存在しないときの画像データAを読み出したが、このステップ140を行わないで、テレビカメラ16で撮影した画像データBと画像データBから所定時間T経過後の画像データCとに基づいて、オブジェクトを認識するようにしてもよい。

【0171】また、本第5の実施形態では2台のテレビカメラを用いてオブジェクトを含む画像を撮影したが、3台以上のテレビカメラを用いても良い。

$$\begin{aligned}\theta &= \tan^{-1} (R/H) & \dots\dots (1)' \\ \theta' &= \tan^{-1} \{R/(H-h)\} & \dots\dots (2)' \\ r &= 2f \sin (\theta/2) & \dots\dots (3)' \\ r' &= 2f \sin (\theta'/2) & \dots\dots (4)' \\ h &= H [1 - \tan \{2 \sin^{-1} (r/2f)\} \\ &\quad / \tan \{2 \sin^{-1} (r'/2f)\}] & \dots\dots (5)' \\ R &= H \tan \{2 \sin^{-1} (r/2f)\} & \dots\dots (6)'\end{aligned}$$

【第6の実施形態】次に、本発明に係る第6の実施形態について説明する。本第6の実施形態では、オブジェクトの音の抽出を行うにあたり、1台のテレビカメラと1枚の鏡とを用いて得たオブジェクトを含む画像データに基づいて、オブジェクトの3次元座標を演算し、オブジェクトの位置を認識する例を示す。なお、本第6の実施形態は前記第5の実施形態と略同一であるので、図13乃至図16において同一の部分には同一の符号を付し、説明を省略する。

【0174】図25に示すように、各テレビカメラ16の1側面には、CCDエリアイメージセンサ16Bの1端面の方向(X方向)と平行で鉛直方向(Z方向)に縦長の鏡74が天井52に固設されている。

【0175】次に、本第6の実施形態の等距離射影型魚眼レンズ16A、CCDエリアイメージセンサ16B及び鏡74の位置、距離及び角度等の諸量を図25、26を参照して説明する。なお、図26は等距離射影型魚眼レンズ16AとCCDエリアイメージセンサ16Bとの距離は微小であるものとして無視したときの上記諸量の詳細を表したものである。

【0176】図25に示すように、CCDエリアイメージセンサ16BのCCD面と同一のXY平面上にある鏡74の上端部の中央を3次元座標の原点O(0, 0, 0)にとる。CCD面の画像中心Hは原点OからY方向に距離hだけ離れており、画像中心Hの3次元座標を(0, h, 0)にとる。なお、対象人物Aの所定部位(例えば頭部)Pの3次元座標を(x, y, z)とし、点Pから出た光は等距離射影型魚眼レンズ16Aで屈折してCCD面上の点Dに結像する。このCCD面上の点Dの2次元座標を(α_D , β_D)とする。また、点Pから出た光で鏡74により反射した光は等距離射影型魚眼レンズ16Aで屈折してCCD面上の点Rに結像する。このCCD面上の点Rの2次元座標を(α_R , β_R)とする。なお、鏡74がないときの仮想のテレビカメラ1※

$$r_D = f \cdot \tan^{-1} [\{ (y-h)^2 + x^2 \}^{1/2} / z]$$

*【0172】また、本第5の実施形態では、等距離射影型魚眼レンズを用いたが、上述したように等立体角射影型魚眼レンズ、立体射影型魚眼レンズや正射影型魚眼レンズを用いても、上記と同様に対象人物Aの頭部の3次元座標を演算することができる。以下に等立体角射影型魚眼レンズを用いたときの式(1)～式(6)に相当する式をそれぞれ次の式(1)'～式(6)'を示す。

【0173】

$$\dots\dots (1)'$$

$$\dots\dots (2)'$$

$$\dots\dots (3)'$$

$$\dots\dots (4)'$$

$$\dots\dots (5)'$$

$$\dots\dots (6)'$$

※7を想定し、CCD面の画像中心H'の3次元座標を(0, -h, 0)にとったときに、点Pから出た光は仮想の等距離射影型魚眼レンズ17Aで屈折して仮想のCCDエリアイメージセンサ17BのCCD面上の点R'に結像するものとし、上述した点Rと仮想した点R'とは鏡74に対して対称であるものとする。また、CCD面上の画像中心Hから点Dまでの距離を r_D 、CCD面上の画像中心Hから点Rまでの距離を r_R とする。

【0177】図26に示したように、点HからZ方向に下ろした垂線上にある任意の点を点Vとし、点H'からZ方向に下ろした垂線上にある任意の点を点V'としたとき、点PHVのなす角を θ_D 、点PH'V'のなす角を θ_R' とする。また、3次元座標(x, y, 0)で表される点を点S、点Sと点Hとの距離を距離 B_R 、点Sと点H'との距離を距離 B_R' 、点Pと点Hとの距離を距離 A_D 、点Pと点H'との距離を距離 A_R' とする。

【0178】次に、本第6の実施形態の作用について説明する。図16に示す抽出位置設定処理におけるステップ160では、対象人物Aを撮影するための1台のテレビカメラ16を選択(例えば距離 r_D が最も小さいテレビカメラを選択)し、対象人物Aの頭部の位置に対応するCCD面上の点D(α_D , β_D)及び点R(α_R , β_R)の各々の2次元座標値を取り込む。

【0179】次のステップ162では、3次元座標を演算する。ここで、図25及び26を参照して上述した諸量について更に説明する。

【0180】角 θ_D 及び θ_R' はそれぞれ、

$$\theta_D = \tan^{-1} (B_D/Q) \\ = \tan^{-1} [\{ (y-h)^2 + x^2 \}^{1/2} / z]$$

$$\theta_R' = \tan^{-1} (B_R'/Q) \\ = \tan^{-1} [\{ (y+h)^2 + x^2 \}^{1/2} / z]$$

により求めることができるので、上記式(3)から距離 r_D 及び r_R は次式により表される。

【0181】

$$r_R = f \cdot \tan^{-1} [\{ (y+h)^2 + x^2 \}^{1/2} / z]$$

ところで、

$$\alpha_D = r_D \cos(\pi - \phi_D) = -r_D \cos \phi_D \quad \dots (17)$$

$$\beta_D = r_D \sin(\pi - \phi_D) = r_D \sin \phi_D \quad \dots (18)$$

$$\alpha_R = r_R \cos \phi_R' \quad (\because \phi_R' = \phi_R) \quad \dots (19)$$

$$\beta_R = r_R \sin \phi_R' \quad (\because \phi_R' = \phi_R) \quad \dots (20)$$

また、

$$\cos \phi_D = (y-h) / \{ (y-h)^2 + x^2 \}^{1/2} \quad \dots (21)$$

$$\sin \phi_D = x / \{ (y-h)^2 + x^2 \}^{1/2} \quad \dots (22)$$

$$\cos \phi_R' = (y+h) / \{ (y+h)^2 + x^2 \}^{1/2} \quad \dots (23)$$

$$\sin \phi_R' = x / \{ (y+h)^2 + x^2 \}^{1/2} \quad \dots (24)$$

であるので、式(17)及び式(21)並びに式(1 * * 8)及び式(22)から、

$$\alpha_D = -f \theta_D (y-h) / \{ (y-h)^2 + x^2 \}^{1/2} \quad \dots (25)$$

$$\beta_D = f \theta_D x / \{ (y-h)^2 + x^2 \}^{1/2} \quad \dots (26)$$

と表すことができる。この2つの式から $f \theta_D$ を消去す※ ※ると、

$$y = h - (\alpha_D / \beta_D) x \quad \dots (27)$$

同様に、

$$\alpha_R = f \theta_R' (y+h) / \{ (y+h)^2 + x^2 \}^{1/2} \quad \dots (28)$$

$$\beta_R = f \theta_R' x / \{ (y+h)^2 + x^2 \}^{1/2} \quad \dots (29)$$

$$y = -h + (\alpha_R / \beta_R) x \quad \dots (30)$$

式(27)及び(30)から

$$x = 2h\beta_D\beta_R / (\alpha_D\beta_R + \alpha_R\beta_D) \quad \dots (31)$$

により3次元座標PのX座標を求めることができる。 ★て、

【0182】次に、式(31)を式(27)に代入し ★

$$y = h (\alpha_R\beta_D - \alpha_D\beta_R) / (\alpha_D\beta_R + \alpha_R\beta_D) \quad \dots (32)$$

により3次元座標PのY座標を求めることができる。 ☆ ☆【0183】また、

$$\begin{aligned} \beta_D &= r_D \sin \phi_D = f \theta_D \sin \phi_D \\ &= f \cdot \tan^{-1} [\{ (y-h)^2 + x^2 \}^{1/2} / z] \cdot \sin \phi_D \end{aligned}$$

この式を変形して、

$$\begin{aligned} z &= \{ (y-h)^2 + x^2 \}^{1/2} / \tan(\beta_D / f \sin \phi_D) \\ &= \{ (y-h)^2 + x^2 \}^{1/2} \\ &\quad / \tan [(\beta_D / f) \times \{ (y-h)^2 + x^2 \}^{1/2} / x] \end{aligned}$$

ところで、式(31)及び式(32)から、

$$\begin{aligned} \{ (y-h)^2 + x^2 \}^{1/2} &= 2h\beta_R (\alpha_D^2 + \beta_D^2)^{1/2} \\ &\quad / (\alpha_D\beta_R + \alpha_R\beta_D) \end{aligned}$$

であるので、

$$\begin{aligned} z &= 2h\beta_R (\alpha_D^2 + \beta_D^2)^{1/2} \\ &\quad / [(\alpha_D\beta_R + \alpha_R\beta_D) \times \tan \{ (\alpha_D^2 + \beta_D^2)^{1/2} / f \}] \\ &\quad \dots (33) \end{aligned}$$

により3次元座標PのZ座標を求めることができる。

【0184】なお、鏡74からのCCD面の画像中心Hまでの距離hは予め定められている。従って、ステップ162では、ROM14Bから式(31)、(32)及び(33)を読み出し、ステップ160で取り込んだCCD面上の点D(α_D , β_D)及び点R(α_R , β_R)の各々の2次元座標値を代入して対象人物Aの頭部の3次元座標P(x, y, z)を演算する。

【0185】このように本第6の実施形態によれば、1台のテレビカメラによって対象人物Aの頭部の3次元座標を演算することができるので、天井52に設置するテ

40 レビカメラの台数を少なくすることができる。

【0186】なお、本第6の実施形態では、天井52に設置された1台のテレビカメラ及び1枚の鏡によりオブジェクトの3次元座標を演算する例を示したが、図24の(G)~(L)に示したように、鏡を壁面に取付けるようにしてもよいし、1台のテレビカメラ及び複数の鏡を使用してもよい。また、湾曲した鏡を使用してもよい。複数の鏡を使用した場合、CCD面にはより多くのオブジェクト像が形成されるので、他の物体(例えば家具や柱等)により死角が生じて、上記のようにして3次元座標を演算することができる。

【0187】〔第7の実施形態〕次に、本発明に係る第7の実施形態について説明する。本第7の実施形態では、オブジェクトの音の抽出を行うにあたり、ボクセルを設定することなくオブジェクトの形状を認識する例を示す。なお、本第7の実施形態は上記第5の実施形態と略同一であるので、図13及び図16の同一の部分には同一の符号を付し、説明を省略する。また、本第7の実施形態では、説明を簡単にするために、図28に示したように対象人物AをテレビカメラA、B、C、Dで捉えた場合を想定する。

【0188】本第7の実施形態における抽出位置演算プロセッサ14は、オブジェクト像を含む歪曲した画像データを、平面化した画像データに変換し、該変換した画像データに基づいて少なくともオブジェクト像の正面、背面、左側面、右側面及び平面の画像データを求め、該求めた画像データを合成してオブジェクトを認識する機能を有する。

【0189】次に、本第7の実施形態の作用について説明する。図27に示す抽出位置演算処理におけるステップ121では、テレビカメラA、B、C、Dにより撮影された画像データを取り込む。このステップ121で取り込んだ画像データの画像は、図29(A)～(D)に示すように、歪曲が生じている。次のステップ123では、これらの歪曲した画像の画像データを、平面化した画像データに変換し、図30(A)～(D)に示すような画像データとする。

【0190】次のステップ125では、平面化した画像データから対象人物Aの前面、背面、左側面、右側面及び平面の画像データを求める。図31(A)～(C)には、このステップ125で求められた対象人物Aの前面、右側面及び平面の画像データをそれぞれ示す。次のステップ127では、ステップ125で求めた前面、背面、左側面、右側面及び平面の画像データを合成する。これにより、オブジェクトの形状を認識することができる。次のステップ134では、前記合成した対象人物Aの画像データに基づいて、図16に示す抽出位置設定処理を上記第5の実施形態と同様に実行する。

【0191】このように本第7の実施形態によれば、ボクセルを設定することなくオブジェクトの形状を認識することができる。

【0192】なお、上記第1～第7の実施形態のテレビカメラ16は可視光のテレビカメラを用いたが、これを例えば、赤外線カメラのように可視光以外の波長域で撮影するようにしてもよい。このようにすれば、照明灯が点灯していないときにもオブジェクトを撮影することができるので、防犯装置や監視装置としても使用することが可能となる。

【0193】また、上記第5～第7の実施形態では、第1の実施形態の音抽出装置10に、広角固定焦点レンズとしての魚眼レンズ16A及びCCDエリアイメージセ

ンサ16Bを備えたテレビカメラ16を適用することにより、オブジェクトの位置、形状を効率的に（速やかに）求める例を示したが、上記第2～第4の実施形態の音抽出装置10に、広角固定焦点レンズとしての魚眼レンズ16A及びCCDエリアイメージセンサ16Bを備えたテレビカメラ16を適用しても同様の効果を得ることができる。

【0194】以上の説明からも明らかなように、本発明は次の技術的態様を含むものである。

10 【0195】前記撮影手段は、3次元空間を構成する部屋の天井に配置されたことを特徴とする請求項1乃至14の何れか1項に記載の音抽出装置。

【0196】前記撮影手段は、可視光以外の波長域で撮影することを特徴とする請求項1乃至14の何れか1項に記載の音抽出装置。

20 【0197】前記形状認識手段は、オブジェクト像を含む歪曲した画像情報を、平面化した画像情報に変換し、該変換した画像情報に基づいて少なくともオブジェクト像の正面、背面、左側面、右側面及び平面の画像情報を求め、該求めた画像情報を合成してオブジェクトを認識することを特徴とする請求項10乃至14の何れか1項に記載の音抽出装置。

30 【0198】人の特徴である高さ、太さ、頭、腕、手、足、顔、目、鼻、口、耳、つま先及び関節に関する情報としての特徴情報の少なくとも1つを予め記憶した記憶手段を更に備え、前記形状認識手段は、前記記憶手段に記憶された特徴情報を読み出し、該特徴情報と撮影手段により撮影された画像情報とに基づいて、オブジェクトが人であることを認識することを特徴とする請求項10乃至14の何れか1項に記載の音抽出装置。

【0199】

【発明の効果】請求項1記載の発明によれば、オブジェクトの位置を認識し且つその位置に基づいてオブジェクトが発する音を周囲の雑音と弁別して抽出することができる、という効果が得られる。

40 【0200】また、請求項2及び16に記載の発明によれば、特にオブジェクトが発する音の指向性が強い場合やオブジェクトの音を発する部分（面）が大きい場合等には、より高い精度の音の抽出を行うことができる、という効果が得られる。

【0201】また、請求項3記載の発明によれば、移動するオブジェクトに対しても、該オブジェクトからの音を抽出することができる、という効果が得られる。

【0202】また、請求項4及び17に記載の発明によれば、複数のオブジェクトに対しても、該複数のオブジェクトの各々からの音を抽出することができる、という効果が得られる。

【0203】また、請求項5及び18に記載の発明によれば、音響環境の状態の変化に応じて精度の高い音の抽出を行うことができる、という効果が得られる。

【0204】また、請求項6及び19に記載の発明によれば、高音域が低い周波数成分よりも相対的に弱くなることを防止することができる、という効果が得られる。

【0205】また、請求項7及び20に記載の発明によれば、オブジェクトの周辺の反射面等の配置状況に応じてより適切な音の抽出を行うことができる、という効果が得られる。

【0206】また、請求項8及び21に記載の発明によれば、音の抽出の精度を低下させることなく、音の抽出に係る処理（抽出手段によるシフト及び抽出処理）の負荷を軽減することができる、という効果が得られる。

【0207】また、請求項9及び22に記載の発明によれば、音抽出装置により音の抽出が可能な領域内にいる人（一人又は複数人）が発した声を音声認識装置へ入力することができる、という効果が得られる。

【0208】また、請求項10に記載の発明によれば、オブジェクトの移動に追従した撮影手段の向きの変更や焦点調整を行うことなく、オブジェクトの位置を速やかに認識することができる、という効果が得られる。また、撮影手段の向きの変更や焦点調整を行うための機械的な作動機構が不要となるので、撮影手段並びに音抽出装置の構造を簡単にすることができると共に、機械的な作動部を減らすことにより耐久性を向上させることができる、という効果も得られる。さらに、認識されたオブジェクトの位置に基づいてオブジェクトが発する音を周囲の雑音と弁別して抽出することができる、という効果も得られる。

【0209】また、請求項11に記載の発明によれば、画像認識手段を構成する形状認識手段及び3次元座標演算手段によって、オブジェクトの3次元座標を速やかに求め、オブジェクトの位置を速やかに認識することができる、という効果が得られる。

【0210】また、請求項12に記載の発明によれば、微小領域はエリアセンサの解像度の限界まで細分化することができるので、画像情報からオブジェクトが占有する微小領域を求めることにより、オブジェクトの形状を細部まで認識することができる、という効果が得られる。

【0211】また、請求項13に記載の発明によれば、複数の撮影手段により撮影された各々異なる画像情報における影の領域を排除することができ、オブジェクトの形状を正確に認識することができる、という効果が得られる。

【0212】また、請求項14に記載の発明によれば、エリアセンサ上に結像された2次元座標を取得し、該取得した複数の2次元座標に基づいてオブジェクトの3次元座標を演算することができる、という効果が得られる。

【0213】また、請求項15に記載の発明によれば、反射手段によりオブジェクト像をエリアセンサ上に結像させることができるので、撮影手段が1つのみであってもオブジェクトの位置を速やかに認識することができ、さ

らにその位置に基づいてオブジェクトが発する音を周囲の雑音と弁別して抽出することができる、という効果が得られる。

【図面の簡単な説明】

【図1】本発明に係る集音の原理を示す模式図である。

【図2】第1～第4の実施形態に係る集音環境を示す模式図である。

【図3】第1、第4の実施形態に係る音抽出装置の概略構成図である。

【図4】第1、第3、第4の実施形態に係る集音位置演算プロセッサで実行される制御ルーチンを示す流図である。

【図5】第1、第2の実施形態に係る音声抽出ボードのプロセッサで実行される制御ルーチンを示す流図である。

【図6】第2の実施形態に係る音抽出装置の概略構成図である。

【図7】第2の実施形態に係る集音位置演算プロセッサで実行される制御ルーチンを示す流図である。

【図8】第3の実施形態に係る音抽出装置の概略構成図である。

【図9】第3の実施形態に係る音声抽出ボードのプロセッサで実行される制御ルーチンを示す流図である。

【図10】屋外での音の抽出に本発明の音抽出装置を適用した場合の構成例である。

【図11】第4の実施形態に係る音の音域による指向性の違いを示す模式図である。

【図12】第4の実施形態に係る音声抽出ボードのプロセッサで実行される制御ルーチンを示す流図である。

【図13】第5～第7の実施形態に係る集音環境を示す模式図である。

【図14】第5、第6の実施形態に係る集音位置演算プロセッサで実行される制御ルーチンを示す流図である。

【図15】オブジェクト分別処理のサブルーチンを示す流図である。

【図16】抽出位置設定処理のサブルーチンを示す流図である。

【図17】オブジェクトを分別する概念を説明する説明図である。

【図18】オブジェクトの高さ等の諸量を説明する説明図である。

【図19】オブジェクトの影の部分とボクセルとの関係を説明する説明図である。

【図20】(A)はテレビカメラAの画像データによるボクセルを示す図であり、(B)はテレビカメラBの画像データによるボクセルを示す図であり、(C)はテレビカメラCの画像データによるボクセルを示す図であり、(D)はテレビカメラDの画像データによるボクセルを示す図である。

【図21】第2次絞り込みで絞り込まれたボクセルの概

念を説明する説明図である。

【図22】第2次絞り込みで絞り込まれたボクセルからダミーモデルに変換する概念を説明する説明図である。

【図23】2つのテレビカメラにより3次元座標を演算するときの諸量を説明する概念図である。

【図24】テレビカメラ又は鏡の各種配置を示す図である。

【図25】第6の実施形態に係る3次元位置認識装置の構成図である。

【図26】第6の実施形態のCCDエリアイメージセンサ等の位置を説明するための説明図である。

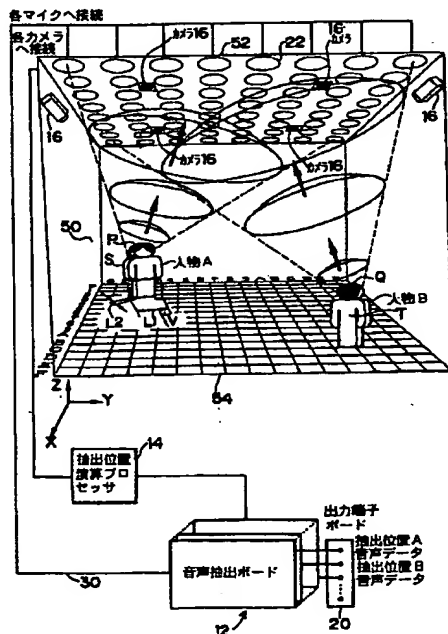
【図27】第7の実施形態に係る集音位置演算プロセッサで実行される制御ルーチンを示す流図である。

【図28】第7の実施形態のオブジェクト及びテレビカメラの配置を示す平面図である。

【図29】(A)はテレビカメラAの画像データの画像を示す図であり、(B)はテレビカメラBの画像データの画像を示す図であり、(C)はテレビカメラCの画像データの画像を示す図であり、(D)はテレビカメラDの画像データの画像を示す図である。

【図30】(A)は歪曲したテレビカメラAの画像データを平面化した画像データに変換したときの画像を示す図であり、(B)は歪曲したテレビカメラBの画像データを平面化した画像データに変換したときの画像を示す図であり、(C)は歪曲したテレビカメラCの画像デー

【図2】



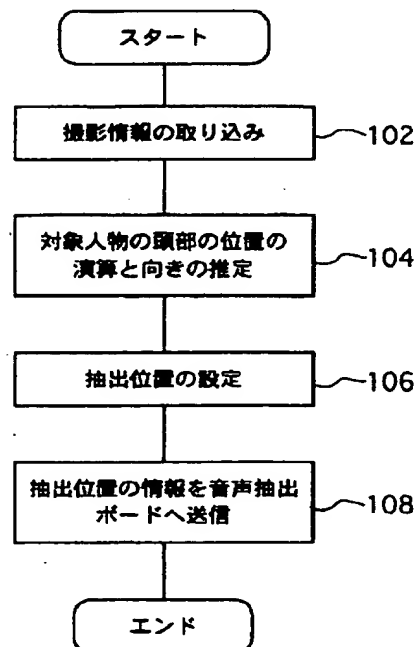
タを平面化した画像データに変換したときの画像を示す図であり、(D)は歪曲したテレビカメラDの画像データを平面化した画像データに変換したときの画像を示す図である。

【図31】(A)は真正面の画像データの画像を示す図であり、(B)は真横の画像データの画像を示す図であり、(C)は真上の画像データの画像を示す図である。

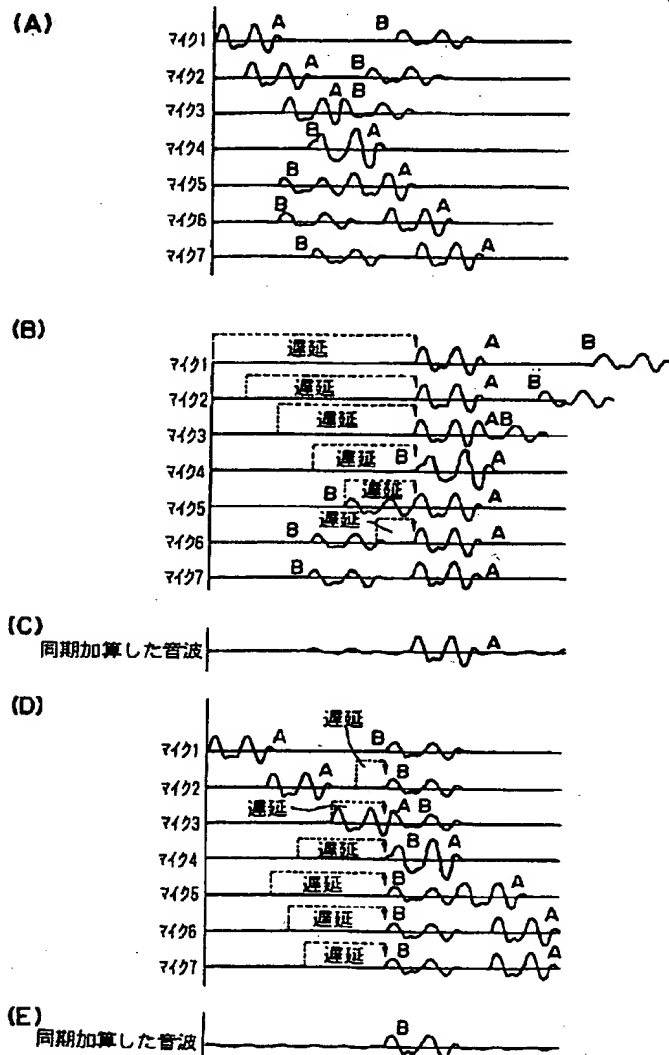
【符号の説明】

- | | |
|-----|-------------------------|
| 10 | 音抽出装置 |
| 12 | 音声抽出ボード |
| 14 | 抽出位置演算プロセッサ |
| 16 | テレビカメラ (撮影手段) |
| 16A | 等距離射影型魚眼レンズ (広角固定焦点レンズ) |
| 16B | CCDエリアイメージセンサ (エリアセンサ) |
| 21 | 音声出力端子 |
| 22 | マイク |
| 32 | 入力バッファメモリ |
| 34 | プロセッサ |
| 44 | 出力バッファメモリ |
| 46 | 加算器 |
| 58 | 温度センサ |
| 60 | 風力計 |
| 62 | 風向計 |
| 74 | 鏡 (反射手段) |

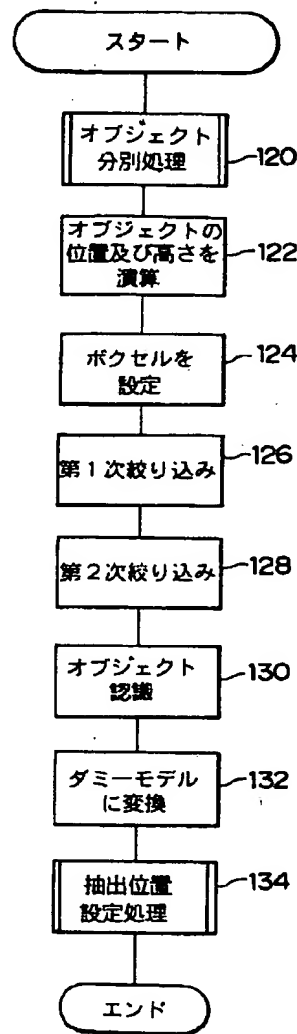
【図4】



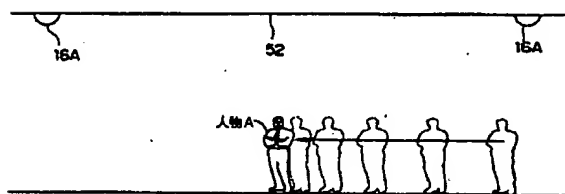
【図1】



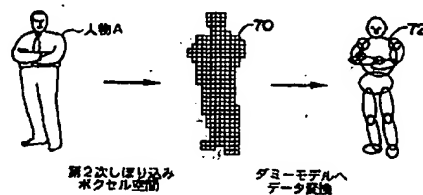
【図14】



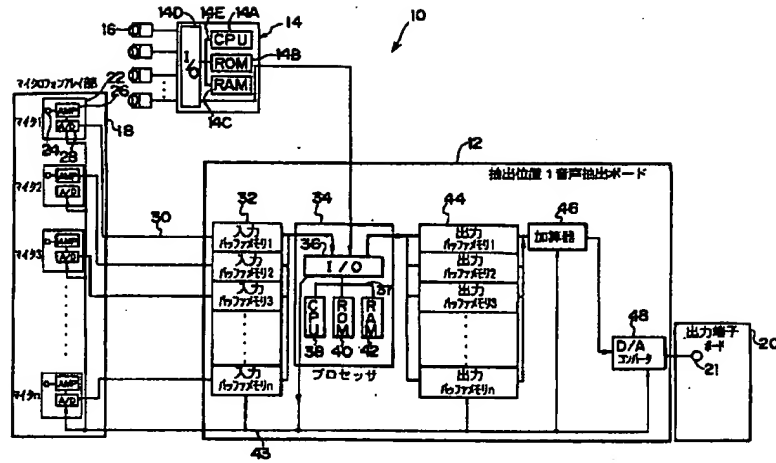
【図1-7】



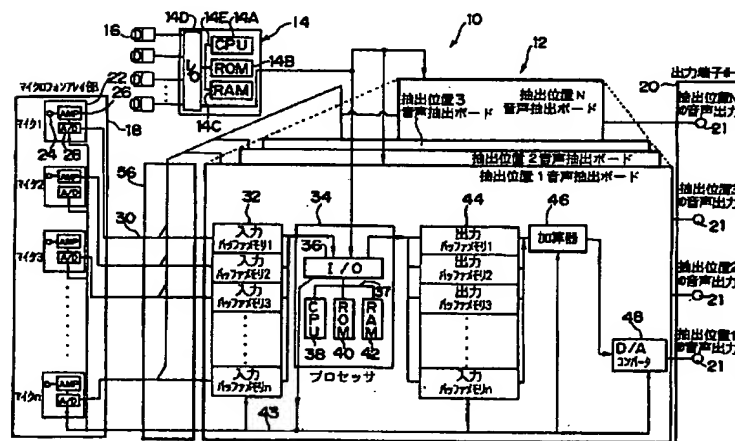
【図22】



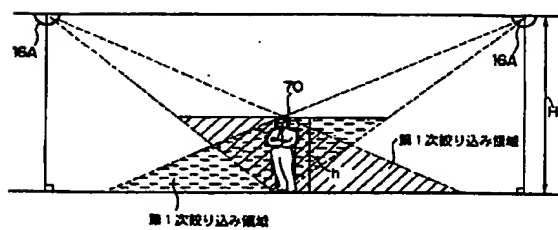
【図3】



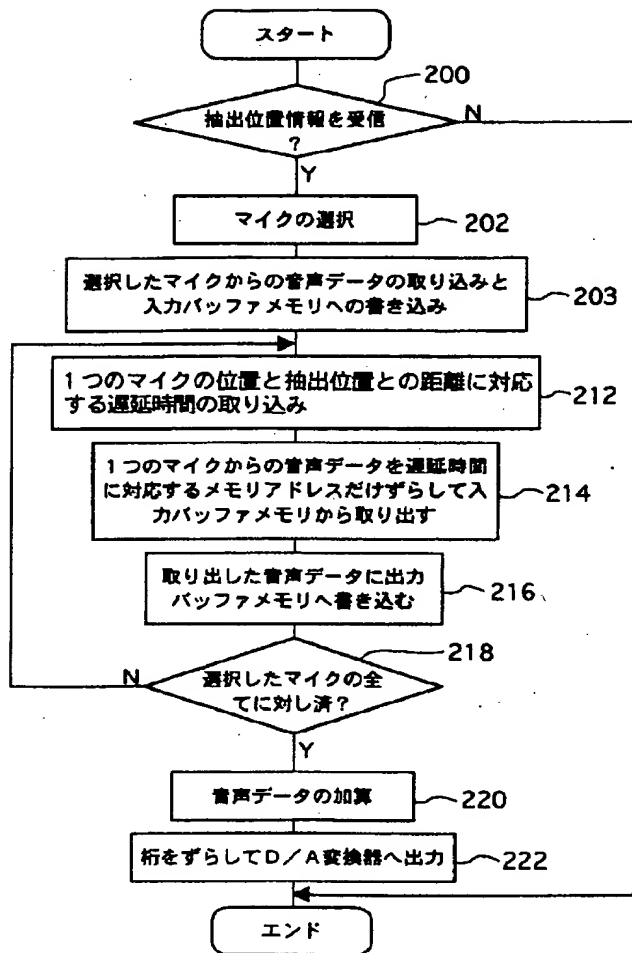
【図6】



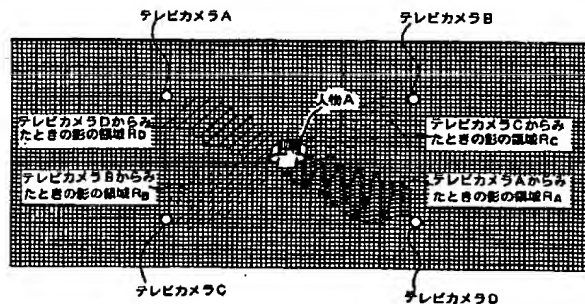
【図21】



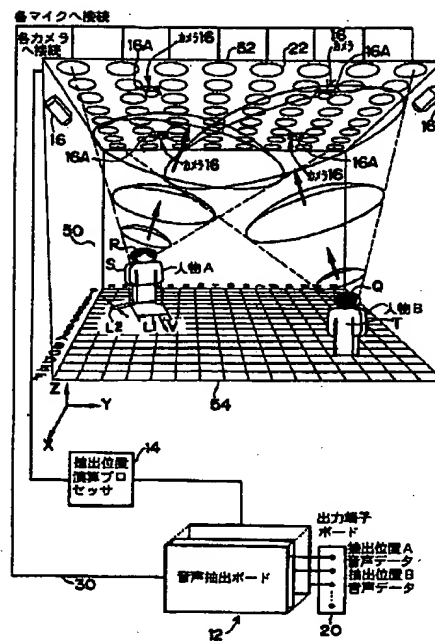
【図5】



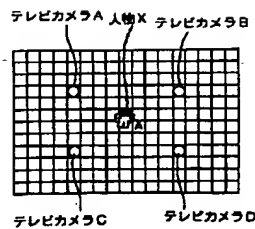
【図19】



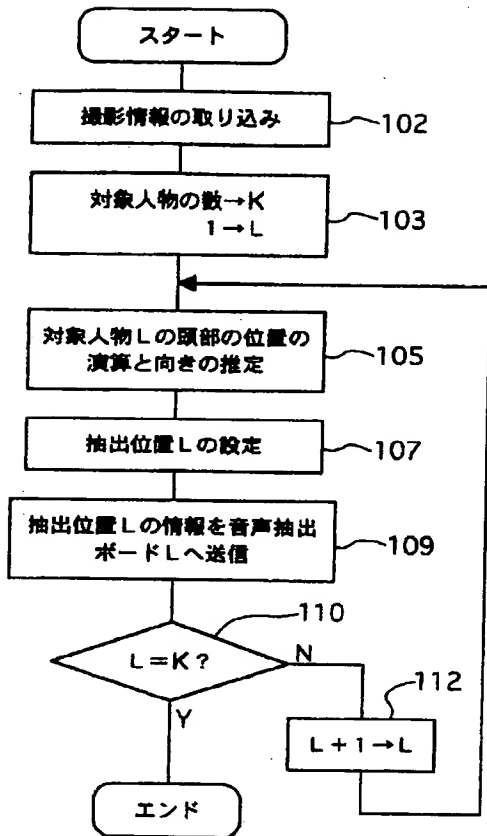
【図13】



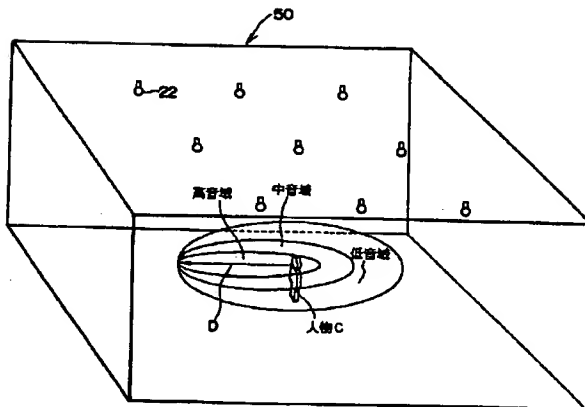
【図28】



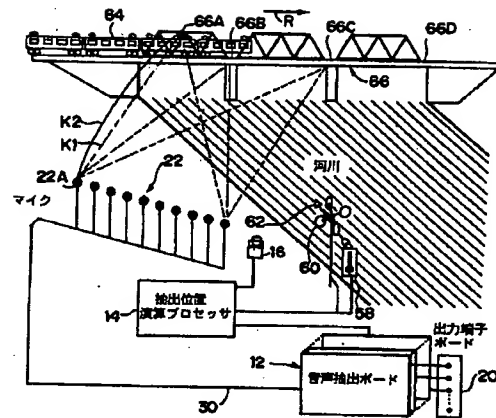
【図7】



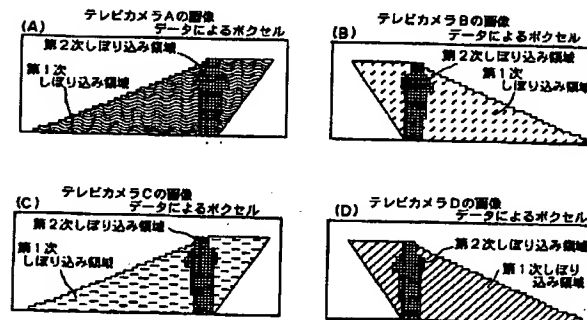
【図11】



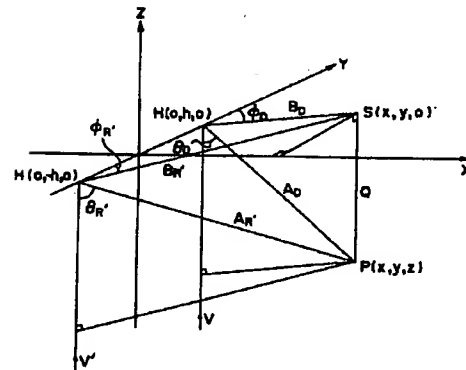
【図10】



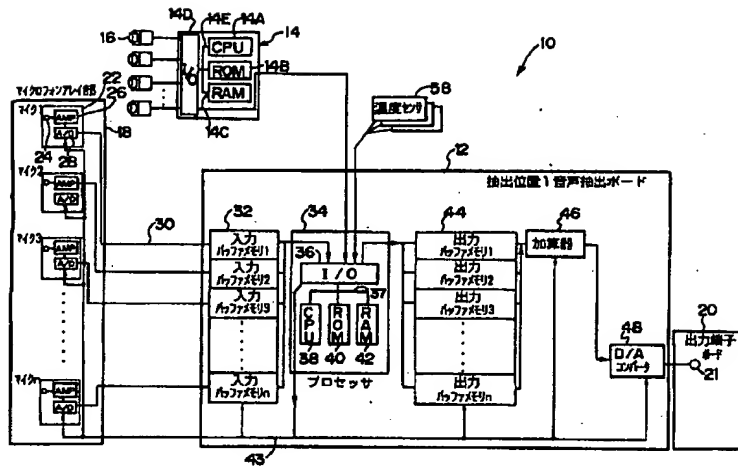
【図20】



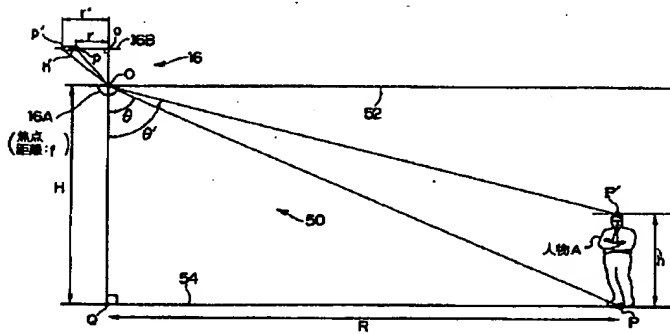
【図26】



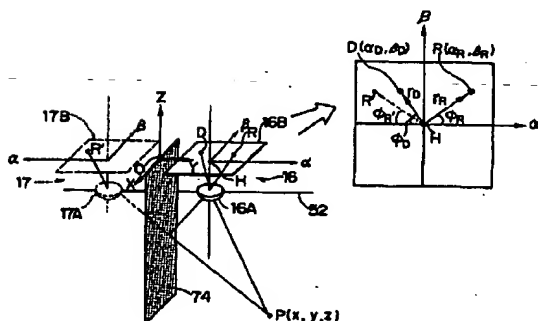
【図8】



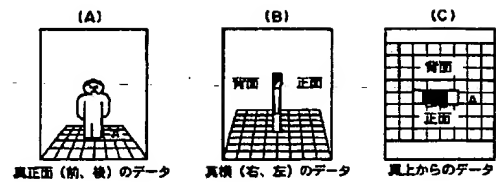
【図18】



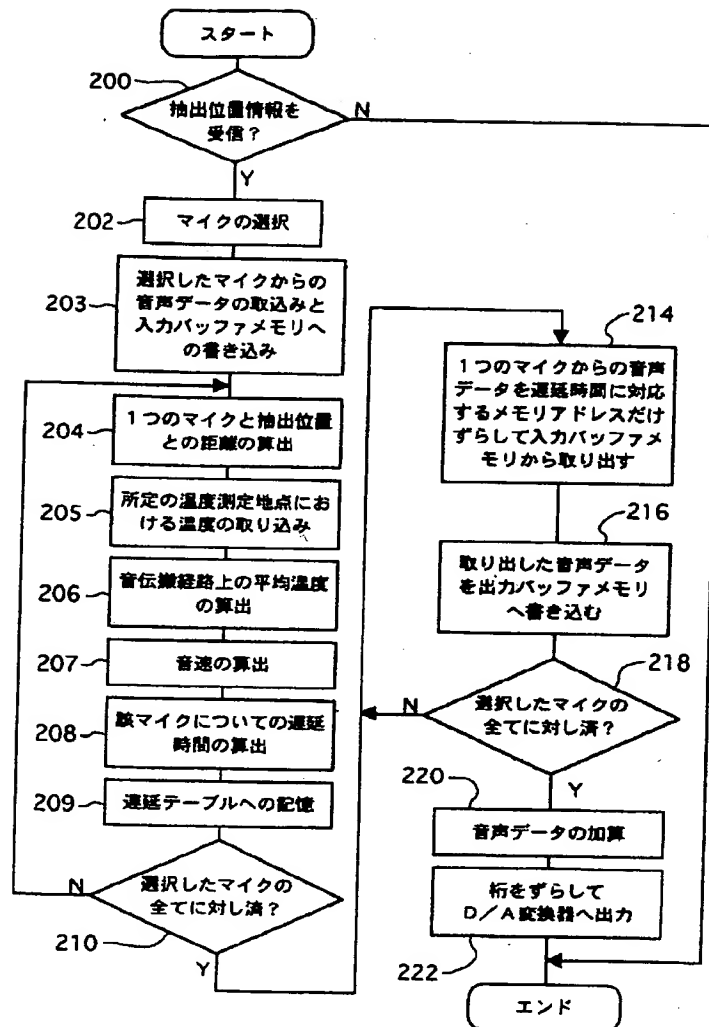
【図25】



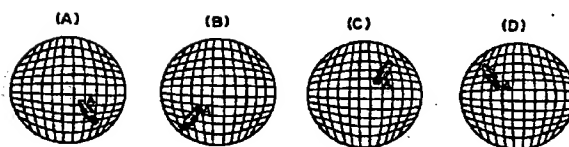
【図31】



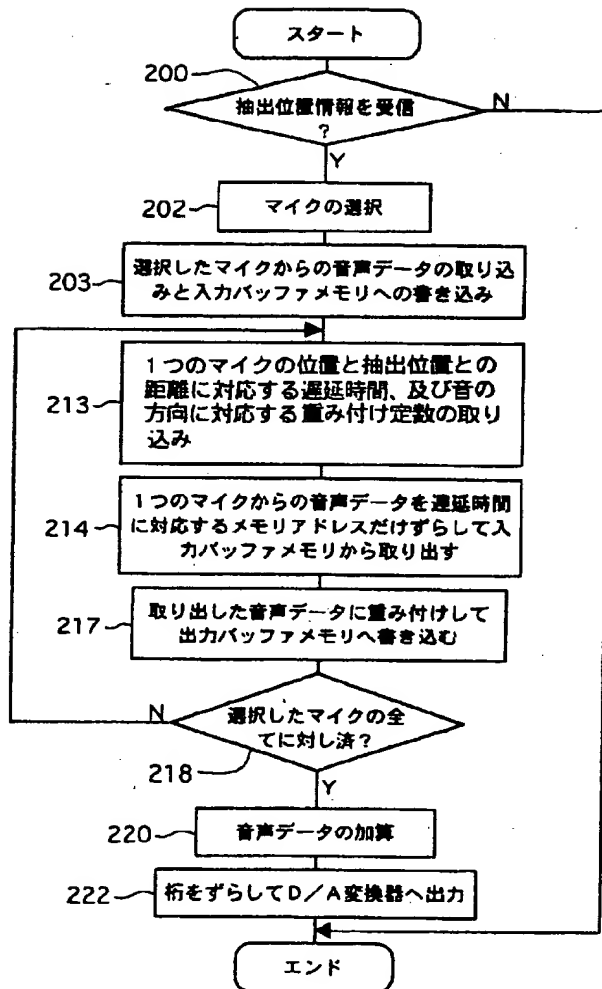
【図9】



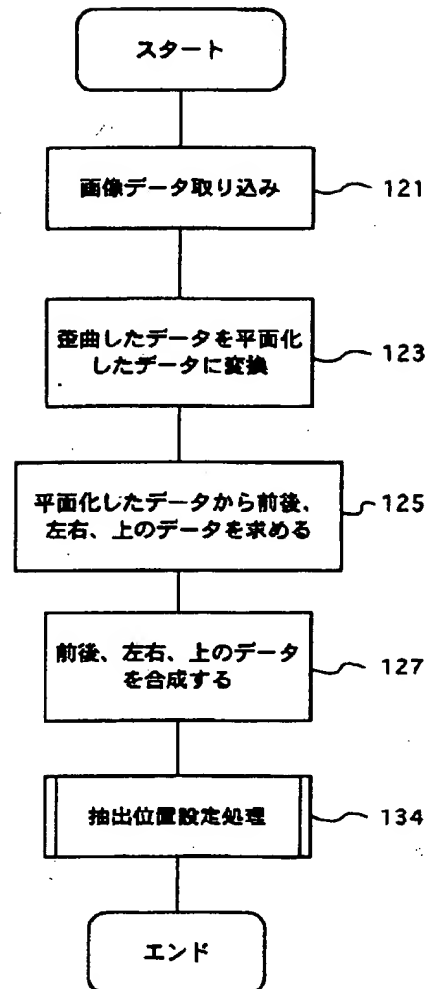
【図29】



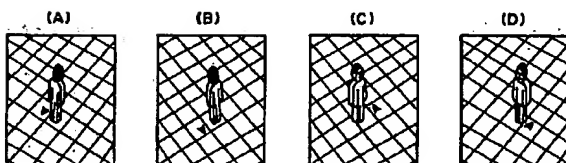
【図12】



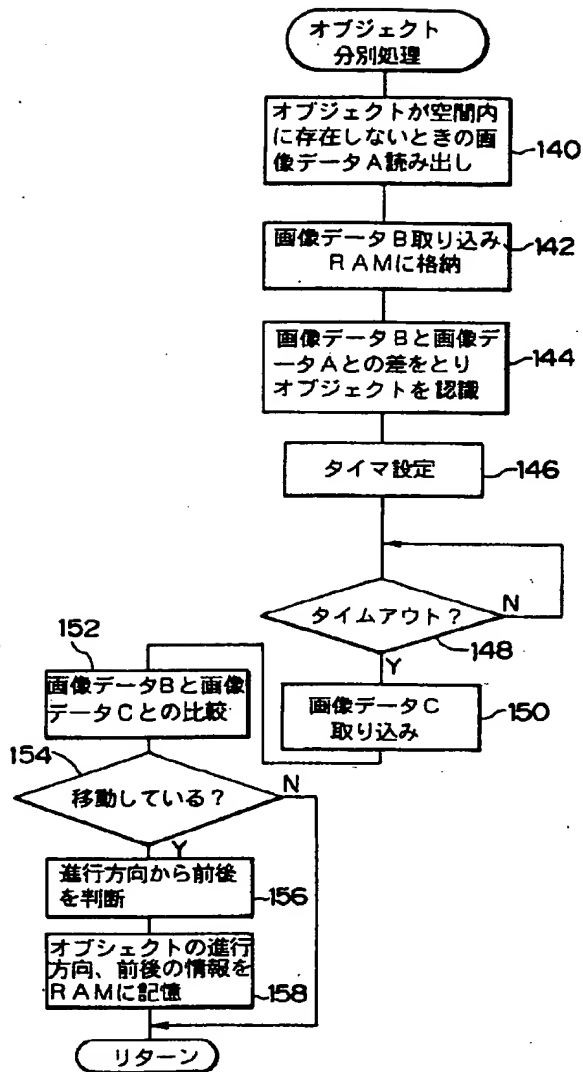
【図27】



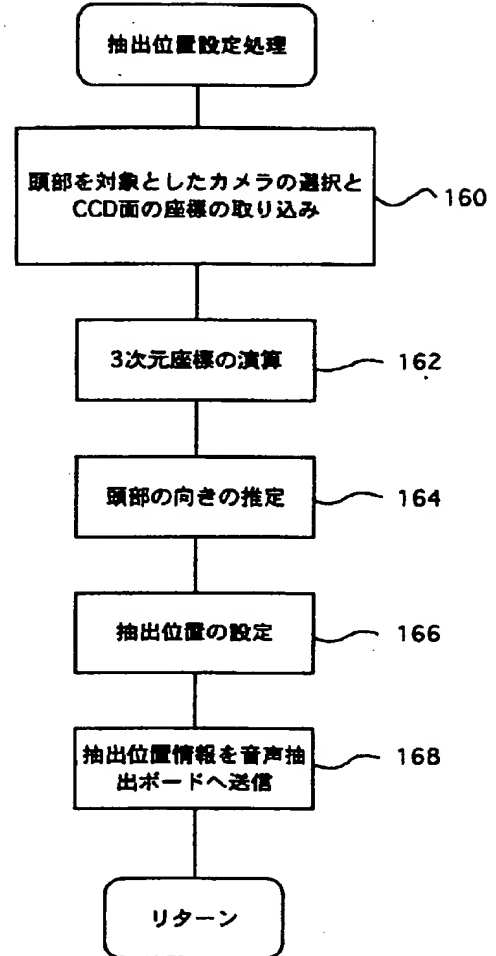
【図30】



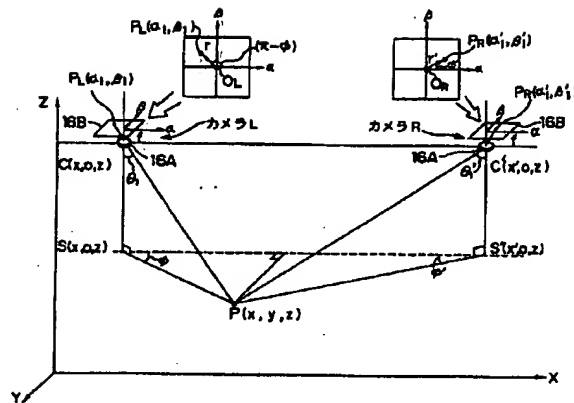
【図15】



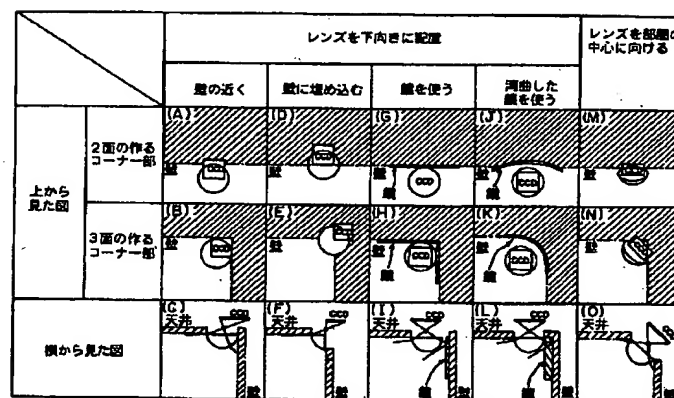
【図16】



【図23】



【図24】



フロントページの続き

(51) Int. Cl. 6

H 0 4 N 7/18

識別記号

庁内整理番号

F I

技術表示箇所

9061-5H

G 0 6 F 15/62

4 1 5

15/70

4 6 0 C

G 1 0 K 15/00

M

B

(72) 発明者 影山 健二

千葉県印旛郡印西町大塚1-5 株式会社
竹中工務店技術研究所内

(72) 発明者 脇中 義孝

千葉県印旛郡印西町大塚1-5 株式会社
竹中工務店技術研究所内

(72) 発明者 中島 立美

千葉県印旛郡印西町大塚1-5 株式会社
竹中工務店技術研究所内

(72) 発明者 山田 祐生

千葉県印旛郡印西町大塚1-5 株式会社
竹中工務店技術研究所内

(72)発明者 海野 健一

千葉県印旛郡印西町大塚1-5 株式会社
竹中工務店技術研究所内

(72)発明者 村井 信義

千葉県印旛郡印西町大塚1-5 株式会社
竹中工務店技術研究所内